

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification:

C12Q 1/68, C12N 15/12

A1

(11) International Publication Number:

WO 92/13103

(43) International Publication Date:

6 August 1992 (06.08.92)

(21) International Application Number: PCT/US92/00376

(22) International Filing Date: 16 January 1992 (16.01.92)

(30) Priority data:

9100963.9
741,94016 January 1991 (16.01.91)
8 August 1991 (08.08.91)GB
US

(71) Applicants: THE JOHNS HOPKINS UNIVERSITY (US); 720 Rutland Avenue, Baltimore, MD 21205 (US); IMPERIAL CHEMICAL INDUSTRIES PLC (GB); Imperial Chemical House, Millbank, London SW1P 1JF (GB); THE UNIVERSITY OF UTAH (US); 421 Wakara Way, Suite 170, Salt Lake City, UT 84108 (US); THE CANCER INSTITUTE (JP/JP); Japanese Foundation for Cancer Research, 37-1, 1-chome Kami-ikebukuro, Toshima-ku, Tokyo 170 (JP).

(72) Inventors: KINZLER, Kenneth, W.; 1348 Holstead Road, Baltimore, MD 21234 (US); VOGELSTEIN, Bert; 3700 Reson Way, Baltimore, MD 21208 (US); ANAND, Rakesh; 62 Grange Way, Sandbach, Cheshire CW11 9ES (GB); HEDGE, Philip, John; 7 Rookery Rise, Winsford, Cheshire CW7 3EA (GB); MARKHAM, Alexander, Fred; 25 Booth Bed Lane, Goostrey, Crewe, Cheshire (GB); ALBERTSEN, Hans; 152 N. M. Street, Salt Lake City, UT 84103 (US); CARLSON, Mary, L.; 2074 E. Sunnyside Avenue, Salt Lake City, UT 84103 (US); GRODEN, Joanna, L.; 629 9th Avenue, Salt Lake City, UT 84103 (US); JOSLYN, Geoff; 426 7th Avenue, Salt Lake City, UT 84103 (US); THLIVELAND, Andrew; 3704 S. 2140 E., Salt Lake City, UT 84109 (US); WHITE,

Raymond, L.; 711 18th Avenue, Salt Lake City, UT 84103 (US); NAKAMURA, Yuzuke; 1-43-3, Matsuyama, Kiyose, Tokyo 204 (JP).

(74) Agents: KACAN, Sarah, A. et al.; Ranner, Riichi, McKie & Becker, 1001 G Street, N.W. 11th Floor, Washington, DC 20001 (US).

(81) Designated States: AT (European patent), AU (European patent), BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH, CI (European patent), CT (OAPI patent), CM (OAPI patent), CS, DE, DE (European patent), DK, DK (European patent), ES, ES (European patent), FI, FR (European patent), GA (OAPI patent), GB, GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU, LU (European patent), MC (European patent), MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL, NL (European patent), NO, PL, RO, RL, SD, SE, SE (European patent), SN (OAPI patent), TD (OAPI patent), TG (OAPI patent).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: INHERITED AND SOMATIC MUTATIONS OF APC GENE IN COLORECTAL CANCER OF HUMANS

(57) Abstract

A human gene termed APC is disclosed. Methods and kits are provided for assessing mutations of the APC gene in human tissues and body samples. APC mutations are found in familial adenomatous polyposis patients as well as in sporadic colorectal cancer patients. APC is expressed in most normal tissues. These results suggest that APC is a tumor suppressor.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	RU	Russian Federation
CC	Congo	KP	Democratic People's Republic of Korea	SU	Soviet Union
CH	Switzerland	KE	Republic of Kenya	SH	Senegal
CI	Côte d'Ivoire	LJ	Luxembourg	SU	Soviet Union
CM	Cameroon	LA	Laos	TU	Turkey
CS	Czechoslovakia	LL	Liechtenstein	TC	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

INHERITED AND SOMATIC MUTATIONS OF APC GENE IN COLORECTAL CANCER OF HUMANS

The U.S. Government has a paid-up license in this invention and the right in limited circumstances to require the patent owner to license others on reasonable terms as provided for by the terms of grants awarded by the National Institutes of Health.

TECHNICAL AREA OF THE INVENTION

The invention relates to the area of cancer diagnostics and therapeutics. More particularly, the invention relates to detection of the germline and somatic alterations of wild-type APC genes. In addition, it relates to therapeutic intervention to restore the function of APC gene product.

BACKGROUND OF THE INVENTION

According to the model of Knudson for tumorigenesis (Cancer Research, Vol. 45, p. 1482, 1985), there are tumor suppressor genes in all normal cells which, when they become non-functional due to mutation, cause neoplastic development. Evidence for this model has been found in the cases of retinoblastoma and colorectal tumors. The implicated suppressor genes in those tumors, RB, p53, DCC and MCC, were found to be deleted or altered in many cases of the tumors studied. (Hansen and Cavenee, Cancer Research, Vol. 47 pp. 5518-5527 (1987); Baker et al., Science, Vol. 244, p. 217 (1989); Fearon et al., Science, Vol. 247, p. 49 (1990); Kinzler et al. Science Vol. 251, p. 1366 (1991).)

In order to fully understand the pathogenesis of tumors, it will be necessary to identify the other suppressor genes that play a role in the tumorigenesis process. Prominent among these is the one(s) presumptively located at 5q21. Cytogenetic (Herrera et al., Am J. Med. Genet., Vol. 25, p. 473 (1986) and linkage (Leppert et al., Science, Vol. 238, p. 1411 (1987); Bodmer et al., Nature, Vol. 328, p. 614 (1987)) studies have shown that this chromosome region harbors the gene

responsible for familial adenomatous polyposis (FAP) and Gardner's Syndrome (GS). FAP is an autosomal-dominant, inherited disease in which affected individuals develop hundreds to thousands of adenomatous polyps, some of which progress to malignancy. GS is a variant of FAP in which desmoid tumors, osteomas and other soft tissue tumors occur together with multiple adenomas of the colon and rectum. A less severe form of polyposis has been identified in which only a few (2-40) polyps develop. This condition also is familial and is linked to the same chromosomal markers as FAP and GS (Leppert et al., New England Journal of Medicine, Vol. 322, pp. 904-908, 1990.) Additionally, this chromosomal region is often deleted from the adenomas (Vogelstein et al., N. Engl. J. Med., Vol. 319, p. 525 (1988)) and carcinomas (Vogelstein et al., N. Engl. J. Med., Vol. 319, p. 525 (1988); Solomon et al., Nature, Vol. 328, p. 616 (1987); Sasaki et al., Cancer Research, Vol. 49, p. 4402 (1989); Delattre et al., Lancet, Vol. 2, p. 353 (1989); and Ashton-Rickardt et al., Oncogene, Vol. 4, p. 1169 (1989)) of patients without FAP (sporadic tumors). Thus, a putative suppressor gene on chromosome 5q21 appears to play a role in the early stages of colorectal neoplasia in both sporadic and familial tumors.

Although the MCC gene has been identified on 5q21 as a candidate suppressor gene, it does not appear to be altered in FAP or GS patients. Thus there is a need in the art for investigations of this chromosomal region to identify genes and to determine if any of such genes are associated with FAP and/or GS and the process of tumorigenesis.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide a method for diagnosing and prognosing a neoplastic tissue of a human.

It is another object of the invention to provide a method of detecting genetic predisposition to cancer.

It is another object of the invention to provide a method of supplying wild-type APC gene function to a cell which has lost said gene function.

It is yet another object of the invention to provide a kit for determination of the nucleotide sequence of APC alleles by the polymerase chain reaction.

It is still another object of the invention to provide nucleic acid probes for detection of mutations in the human APC gene.

It is still another object of the invention to provide a cDNA molecule encoding the APC gene product.

It is yet another object of the invention to provide a preparation of the human APC protein.

It is another object of the invention to provide a method of screening for genetic predisposition to cancer.

It is an object of the invention to provide methods of testing therapeutic agents for the ability to suppress neoplasia.

It is still another object of the invention to provide animals carrying mutant APC alleles.

These and other objects of the invention are provided by one or more of the embodiments which are described below. In one embodiment of the present invention a method of diagnosing or prognosing a neoplastic tissue of a human is provided comprising: detecting somatic alteration of wild-type APC genes or their expression products in a sporadic colorectal cancer tissue, said alteration indicating neoplasia of the tissue.

In yet another embodiment a method is provided of detecting genetic predisposition to cancer in a human including familial adenomatous polyposis (FAP) and Gardner's Syndrome (GS), comprising: isolating a human sample selected from the group consisting of blood and fetal tissue; detecting alteration of wild-type APC gene coding sequences or their expression products from the sample, said alteration indicating genetic predisposition to cancer.

In another embodiment of the present invention a method is provided for supplying wild-type APC gene function to a cell which has lost said gene function by virtue of a mutation in the APC gene, comprising: introducing a wild-type APC gene into a cell which has lost said gene function such that said wild-type gene is expressed in the cell.

In another embodiment a method of supplying wild-type APC gene function to a cell is provided comprising: introducing a portion of a wild-type APC gene into a cell which has lost said gene function such

that said portion is expressed in the cell, said portion encoding a part of the APC protein which is required for non-neoplastic growth of said cell. APC protein can also be applied to cells or administered to animals to remediate for mutant APC genes. Synthetic peptides or drugs can also be used to mimic APC function in cells which have altered APC expression.

In yet another embodiment a pair of single stranded primers is provided for determination of the nucleotide sequence of the APC gene by polymerase chain reaction. The sequence of said pair of single stranded DNA primers is derived from chromosome 5q band 21, said pair of primers allowing synthesis of APC gene coding sequences.

In still another embodiment of the invention a nucleic acid probe is provided which is complementary to human wild-type APC gene coding sequences and which can form mismatches with mutant APC genes, thereby allowing their detection by enzymatic or chemical cleavage or by shifts in electrophoretic mobility.

In another embodiment of the invention a method is provided for detecting the presence of a neoplastic tissue in a human. The method comprises isolating a body sample from a human; detecting in said sample alteration of a wild-type APC gene sequence or wild-type APC expression product, said alteration indicating the presence of a neoplastic tissue in the human.

In still another embodiment a cDNA molecule is provided which comprises the coding sequence of the APC gene.

In even another embodiment a preparation of the human APC protein is provided which is substantially free of other human proteins. The amino acid sequence of the protein is shown in Figure 3 or 7.

In yet another embodiment of the invention a method is provided for screening for genetic predisposition to cancer, including familial adenomatous polyposis (FAP) and Gardner's Syndrome (GS), in a human. The method comprises: detecting among kindred persons the presence of a DNA polymorphism which is linked to a mutant APC allele in an individual having a genetic predisposition to cancer, said kindred being genetically related to the individual, the presence of said polymorphism suggesting a predisposition to cancer.

In another embodiment of the invention a method of testing therapeutic agents for the ability to suppress a neoplastically transformed phenotype is provided. The method comprises: applying a test substance to a cultured epithelial cell which carries a mutation in an APC allele; and determining whether said test substance suppresses the neoplastically transformed phenotype of the cell.

In another embodiment of the invention a method of testing therapeutic agents for the ability to suppress a neoplastically transformed phenotype is provided. The method comprises: administering a test substance to an animal which carries a mutant APC allele; and determining whether said test substance prevents or suppresses the growth of tumors.

In still other embodiments of the invention transgenic animals are provided. The animals carry a mutant APC allele from a second animal species or have been genetically engineered to contain an insertion mutation which disrupts an APC allele.

The present invention provides the art with the information that the APC gene, a heretofore unknown gene is, in fact, a target of mutational alterations on chromosome 5q21 and that these alterations are associated with the process of tumorigenesis. This information allows highly specific assays to be performed to assess the neoplastic status of a particular tissue or the predisposition to cancer of an individual. This invention has applicability to Familial Adenomatous Polyposis, sporadic colorectal cancers, Gardner's Syndrome, as well as the less severe familial polyposis discussed above.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A shows an overview of yeast artificial chromosome (YAC) contigs. Genetic distances between selected RFLP markers from within the contigs are shown in centiMorgans.

Figure 1B shows a detailed map of the three central contigs. The position of the six identified genes from within the FAP region is shown: the 5' and 3' ends of the transcripts from these genes have in general not yet been isolated, as indicated by the string of dots surrounding the bars denoting the genes' positions. Selected restriction

endonuclease recognition sites are indicated: B, BssH2; S, SsII; M, MluI; N, NruI.

Figure 2 shows the sequence of TB1 and TB2 genes. The cDNA sequence of the TB1 gene was determined from the analysis of 11 cDNA clones derived from normal colon and liver, as described in the text. A total of 2314 bp were contained within the overlapping cDNA clones, defining an ORF of 424 amino acids beginning at nucleotide 1. Only the predicted amino acids from the ORF are shown. The carboxy-terminal end of the ORF has apparently been identified, but the 5' end of the TB1 transcript has not yet been precisely determined.

The cDNA sequence of the TB2 gene was determined from the YS-39 clone derived as described in the text. This clone consisted of 2300 bp and defined an ORF of 185 amino acids beginning at nucleotide 1. Only the predicted amino acids are shown. The carboxy terminal end of the ORF has apparently been identified, but the 5' end of the TB2 transcript has not been precisely determined.

Figure 3 shows the sequence of the APC gene product. The cDNA sequence was determined through the analysis of 87 cDNA clones derived from normal colon, liver, and brain. A total of 8973 bp were contained within overlapping cDNA clones, defining an ORF of 2842 amino acids. In frame stop codons surrounded this ORF, as described in the text, suggesting that the entire APC gene product was represented in the ORF illustrated. Only the predicted amino acids are shown.

Figure 4 shows the local similarity between human APC and ral2 of yeast. Local similarity among the APC and MCC genes and the m3 muscarinic acetylcholine receptor is shown. The region of the mAChR shown corresponds to that responsible for coupling the receptor to G proteins. The connecting lines indicate identities; dots indicate related amino acids residues.

Figure 5 shows the genomic map of the 1200 kb NotI fragment at the FAP locus. The NotI fragment is shown as a bold line. Relevant parts of the deletion chromosomes from patients 3214 and 3824 are shown as stippled lines. Probes used to characterize the NotI fragment and the deletions, and three YACs from which subclones were obtained, are shown below the restriction map. The chimeric end of YAC

183H12 is indicated by a dotted line. The orientation and approximate position of MCC are indicated above the map.

Figure 6 shows the DNA sequence and predicted amino acid sequence of DP1 (TB2). The nucleotide numbering begins at the most 5' nucleotide isolated. A proposed initiation methionine (base 77) is indicated in bold type. The entire coding sequence is presented.

Figure 7 shows the cDNA and predicted amino acid sequence of DP2.5 (APC). The nucleotide numbering begins at the proposed initiation methionine. The nucleotides and amino acids of the alternatively spliced-exon (exon 9; nucleotide positions 934-1236) are presented in lower case letters. At the 3' end, a poly(A) addition signal occurs at 8530, and one cDNA clone has a poly(A) at 9563. Other cDNA clones extend beyond 9563, however, and their consensus sequence is included here.

Figure 8 shows the arrangement of exons in DP2.5 (APC). (A) Exon 9 corresponds to nucleotides 933-1312; exon 9a corresponds to nucleotides 1236-1312. The stop codon in the cDNA is at nucleotide 8535. (B) Partial intronic sequence surrounding each exon is shown.

DETAILED DESCRIPTION

It is a discovery of the present invention that mutational events associated with tumorigenesis occur in a previously unknown gene on chromosome 5q named here the APC (Adenomatous Polyposis Coli) gene. Although it was previously known that deletion of alleles on chromosome 5q were common in certain types of cancers, it was not known that a target gene of these deletions was the APC gene. Further it was not known that other types of mutational events in the APC gene are also associated with cancers. The mutations of the APC gene can involve gross rearrangements, such as insertions and deletions. Point mutations have also been observed.

According to the diagnostic and prognostic method of the present invention, alteration of the wild-type APC gene is detected. "Alteration of a wild-type gene" according to the present invention encompasses all forms of mutations — including deletions. The alteration may be due to either rearrangements such as insertions, inversions, and deletions, or to point mutations. Deletions may be of the

entire gene or only a portion of the gene. Somatic mutations are those which occur only in certain tissues, e.g., in the tumor tissue, and are not inherited in the germline. Germline mutations can be found in any of a body's tissues. If only a single allele is somatically mutated, an early neoplastic state is indicated. However, if both alleles are mutated then a late neoplastic state is indicated. The finding of APC mutations thus provides both diagnostic and prognostic information. An APC allele which is not deleted (e.g., that on the sister chromosome to a chromosome carrying an APC deletion) can be screened for other mutations, such as insertions, small deletions, and point mutations. It is believed that many mutations found in tumor tissues will be those leading to decreased expression of the APC gene product. However, mutations leading to non-functional gene products would also lead to a cancerous state. Point mutational events may occur in regulatory regions, such as in the promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to loss of expression of the APC gene product.

In order to detect the alteration of the wild-type APC gene in a tissue, it is helpful to isolate the tissue free from surrounding normal tissues. Means for enriching a tissue preparation for tumor cells are known in the art. For example, the tissue may be isolated from paraffin or cryostat sections. Cancer cells may also be separated from normal cells by flow cytometry. These as well as other techniques for separating tumor from normal cells are well known in the art. If the tumor tissue is highly contaminated with normal cells, detection of mutations is more difficult.

Detection of point mutations may be accomplished by molecular cloning of the APC allele (or alleles) and sequencing that allele(s) using techniques well known in the art. Alternatively, the polymerase chain reaction (PCR) can be used to amplify gene sequences directly from a genomic DNA preparation from the tumor tissue. The DNA sequence of the amplified sequences can then be determined. The polymerase chain reaction itself is well known in the art. See, e.g., Saiki et al., Science, Vol. 239, p. 487, 1988; U.S. 4,683,203; and U.S. 4,663,195.

Specific primers which can be used in order to amplify the gene will be discussed in more detail below. The ligase chain reaction, which is known in the art, can also be used to amplify APC sequences. See Wu et al., Genomics, Vol. 4, pp. 560-569 (1989). In addition, a technique known as allele specific PCR can be used. (See Ruano and Kidd, Nucleic Acids Research, Vol. 17, p. 8392, 1989.) According to this technique, primers are used which hybridize at their 3' ends to a particular APC mutation. If the particular APC mutation is not present, an amplification product is not observed. Amplification Refractory Mutation System (ARMS) can also be used as disclosed in European Patent Application Publication No. 0332435 and in Newton et al., Nucleic Acids Research, Vol. 17, p.7, 1989. Insertions and deletions of genes can also be detected by cloning, sequencing and amplification. In addition, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding marker genes can be used to score alteration of an allele or an insertion in a polymorphic fragment. Such a method is particularly useful for screening among kindred persons of an affected individual for the presence of the APC mutation found in that individual. Single stranded conformation polymorphism (SSCP) analysis can also be used to detect base change variants of an allele. (Orita et al., Proc. Natl. Acad. Sci. USA Vol. 86, pp. 2766-2770, 1989, and Genomics, Vol. 5, pp. 874-879, 1989.) Other techniques for detecting insertions and deletions as are known in the art can be used.

Alteration of wild-type genes can also be detected on the basis of the alteration of a wild-type expression product of the gene. Such expression products include both the APC mRNA as well as the APC protein product. The sequences of these products are shown in Figures 3 and 7. Point mutations may be detected by amplifying and sequencing the mRNA or via molecular cloning of cDNA made from the mRNA. The sequence of the cloned cDNA can be determined using DNA sequencing techniques which are well known in the art. The cDNA can also be sequenced via the polymerase chain reaction (PCR) which will be discussed in more detail below.

Mismatches, according to the present invention are hybridized nucleic acid duplexes which are not 100% homologous. The lack of

total homology may be due to deletions, insertions, inversions, substitutions or frameshift mutations. Mismatch detection can be used to detect point mutations in the gene or its mRNA product. While these techniques are less sensitive than sequencing, they are simpler to perform on a large number of tumor samples. An example of a mismatch cleavage technique is the RNase protection method, which is described in detail in Winter et al., Proc. Natl. Acad. Sci. USA, Vol. 82, p. 7575, 1985 and Meyers et al., Science, Vol. 230, p. 1242, 1985. In the practice of the present invention the method involves the use of a labeled riboprobe which is complementary to the human wild-type APC gene coding sequence. The riboprobe and either mRNA or DNA isolated from the tumor tissue are annealed (hybridized) together and subsequently digested with the enzyme RNase A which is able to detect some mismatches in a duplex RNA structure. If a mismatch is detected by RNase A, it cleaves at the site of the mismatch. Thus, when the annealed RNA preparation is separated on an electrophoretic gel matrix, if a mismatch has been detected and cleaved by RNase A, an RNA product will be seen which is smaller than the full-length duplex RNA for the riboprobe and the mRNA or DNA. The riboprobe need not be the full length of the APC mRNA or gene but can be a segment of either. If the riboprobe comprises only a segment of the APC mRNA or gene it will be desirable to use a number of these probes to screen the whole mRNA sequence for mismatches.

In similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. See, e.g., Cotton et al., Proc. Natl. Acad. Sci. USA, Vol. 85, 4397, 1988; and Shenk et al., Proc. Natl. Acad. Sci. USA, Vol. 72, p. 989, 1975. Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to matched duplexes. See, e.g., Carilello, Human Genetics, Vol. 42, p. 726, 1988. With either riboprobes or DNA probes, the cellular mRNA or DNA which might contain a mutation can be amplified using PCR (see below) before hybridization. Changes in DNA of the APC gene can also be detected using Southern hybridization, especially if the changes are gross rearrangements, such as deletions and insertions.

DNA sequences of the APC gene which have been amplified by use of polymerase chain reaction may also be screened using allele-specific probes. These probes are nucleic acid oligomers, each of which contains a region of the APC gene sequence harboring a known mutation. For example, one oligomer may be about 30 nucleotides in length, corresponding to a portion of the APC gene sequence. By use of a battery of such allele-specific probes, PCR amplification products can be screened to identify the presence of a previously identified mutation in the APC gene. Hybridization of allele-specific probes with amplified APC sequences can be performed, for example, on a nylon filter. Hybridization to a particular probe under stringent hybridization conditions indicates the presence of the same mutation in the tumor tissue as in the allele-specific probe.

Alteration of APC mRNA expression can be detected by any technique known in the art. These include Northern blot analysis, PCR amplification and RNase protection. Diminished mRNA expression indicates an alteration of the wild-type APC gene.

Alteration of wild-type APC genes can also be detected by screening for alteration of wild-type APC protein. For example, monoclonal antibodies immunoreactive with APC can be used to screen a tissue. Lack of cognate antigen would indicate an APC mutation. Antibodies specific for products of mutant alleles could also be used to detect mutant APC gene product. Such immunological assays can be done in any convenient format known in the art. These include Western blots, immunohistochemical assays and ELISA assays. Any means for detecting an altered APC protein can be used to detect alteration of wild-type APC genes. Functional assays can be used, such as protein binding determinations. For example, it is believed that APC protein oligomerizes to itself and/or MCC protein or binds to a G protein. Thus, an assay for the ability to bind to wild type APC or MCC protein or that G protein can be employed. In addition, assays can be used which detect APC biochemical function. It is believed that APC is involved in phospholipid metabolism. Thus, assaying the enzymatic products of the involved phospholipid metabolic pathway can be used to

determine APC activity. Finding a mutant APC gene product indicates alteration of a wild-type APC gene.

Mutant APC genes or gene products can also be detected in other human body samples, such as, serum, stool, urine and sputum. The same techniques discussed above for detection of mutant APC genes or gene products in tissues can be applied to other body samples. Cancer cells are sloughed off from tumors and appear in such body samples. In addition, the APC gene product itself may be secreted into the extracellular space and found in these body samples even in the absence of cancer cells. By screening such body samples, a simple early diagnosis can be achieved for many types of cancers. In addition, the progress of chemotherapy or radiotherapy can be monitored more easily by testing such body samples for mutant APC genes or gene products.

The methods of diagnosis of the present invention are applicable to any tumor in which APC has a role in tumorigenesis. Deletions of chromosome arm 5q have been observed in tumors of lung, breast, colon, rectum, bladder, liver, sarcomas, stomach and prostate, as well as in leukemias and lymphomas. Thus these are likely to be tumors in which APC has a role. The diagnostic method of the present invention is useful for clinicians so that they can decide upon an appropriate course of treatment. For example, a tumor displaying alteration of both APC alleles might suggest a more aggressive therapeutic regimen than a tumor displaying alteration of only one APC allele.

The primer pairs of the present invention are useful for determination of the nucleotide sequence of a particular APC allele using the polymerase chain reaction. The pairs of single stranded DNA primers can be annealed to sequences within or surrounding the APC gene on chromosome 5q in order to prime amplifying DNA synthesis of the APC gene itself. A complete set of these primers allows synthesis of all of the nucleotides of the APC gene coding sequences, i.e., the exons. The set of primers preferably allows synthesis of both intron and exon sequences. Allele specific primers can also be used. Such primers anneal only to particular APC mutant alleles, and thus will only amplify a product in the presence of the mutant allele as a template.

In order to facilitate subsequent cloning of amplified sequences, primers may have restriction enzyme site sequences appended to their 3' ends. Thus, all nucleotides of the primers are derived from APC sequences or sequences adjacent to APC except the few nucleotides necessary to form a restriction enzyme site. Such enzymes and sites are well known in the art. The primers themselves can be synthesized using techniques which are well known in the art. Generally, the primers can be made using oligonucleotide synthesizing machines which are commercially available. Given the sequence of the APC open reading frame shown in Figure 7, design of particular primers is well within the skill of the art.

The nucleic acid probes provided by the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA and in the RNase protection method for detecting point mutations already discussed above. The probes can be used to detect PCR amplification products. They may also be used to detect mismatches with the APC gene or mRNA using other techniques. Mismatches can be detected using either enzymes (e.g., S1 nuclease), chemicals (e.g., hydroxylamine or osmium tetroxide and piperidine), or changes in electrophoretic mobility of mismatched hybrids as compared to totally matched hybrids. These techniques are known in the art. See, Cotton, *supra*, Shenk, *supra*, Myers, *supra*, Winter, *supra*, and Novack et al., Proc. Natl. Acad. Sci. USA, Vol. 83, p. 386, 1986. Generally, the probes are complementary to APC gene coding sequences, although probes to certain introns are also contemplated. An entire battery of nucleic acid probes is used to compose a kit for detecting alteration of wild-type APC genes. The kit allows for hybridization to the entire APC gene. The probes may overlap with each other or be contiguous.

If a riboprobe is used to detect mismatches with mRNA, it is complementary to the mRNA of the human wild-type APC gene. The riboprobe thus is an anti-sense probe in that it does not code for the APC protein because it is of the opposite polarity to the sense strand. The riboprobe generally will be labeled with a radioactive, colorimetric, or fluorometric material, which can be accomplished by

any means known in the art. If the riboprobe is used to detect mismatches with DNA it can be of either polarity, sense or anti-sense. Similarly, DNA probes also may be used to detect mismatches.

Nucleic acid probes may also be complementary to mutant alleles of the APC gene. These are useful to detect similar mutations in other patients on the basis of hybridization rather than mismatches. These are discussed above and referred to as allele-specific probes. As mentioned above, the APC probes can also be used in Southern hybridizations to genomic DNA to detect gross chromosomal changes such as deletions and insertions. The probes can also be used to select cDNA clones of APC genes from tumor and normal tissues. In addition, the probes can be used to detect APC mRNA in tissues to determine if expression is diminished as a result of alteration of wild-type APC genes. Provided with the APC coding sequence shown in Figure 7 (SEQ ID NO: 1), design of particular probes is well within the skill of the ordinary artisan.

According to the present invention a method is also provided of supplying wild-type APC function to a cell which carries mutant APC alleles. Supplying such function should suppress neoplastic growth of the recipient cells. The wild-type APC gene or a part of the gene may be introduced into the cell in a vector such that the gene remains extrachromosomal. In such a situation the gene will be expressed by the cell from the extrachromosomal location. If a gene portion is introduced and expressed in a cell carrying a mutant APC allele, the gene portion should encode a part of the APC protein which is required for non-neoplastic growth of the cell. More preferred is the situation where the wild-type APC gene or a part of it is introduced into the mutant cell in such a way that it recombines with the endogenous mutant APC gene present in the cell. Such recombination requires a double recombination event which results in the correction of the APC gene mutation. Vectors for introduction of genes both for recombination and for extrachromosomal maintenance are known in the art and any suitable vector may be used. Methods for introducing DNA into cells such as electroporation, calcium phosphate co-precipitation and viral transduction are known in the art and the choice of method is

within the competence of the routineer. Cells transformed with the wild-type APC gene can be used as model systems to study cancer remission and drug treatments which promote such remission.

Similarly, cells and animals which carry a mutant APC allele can be used as model systems to study and test for substances which have potential as therapeutic agents. The cells are typically cultured epithelial cells. These may be isolated from individuals with APC mutations, either somatic or germline. Alternatively, the cell line can be engineered to carry the mutation in the APC allele. After a test substance is applied to the cells, the neoplastically transformed phenotype of the cell will be determined. Any trait of neoplastically transformed cells can be assessed, including anchorage-independent growth, tumorigenicity in nude mice, invasiveness of cells, and growth factor dependence. Assays for each of these traits are known in the art.

Animals for testing therapeutic agents can be selected after mutagenesis of whole animals or after treatment of germline cells or zygotes. Such treatments include insertion of mutant APC alleles, usually from a second animal species, as well as insertion of disrupted homologous genes. Alternatively, the endogenous APC gene(s) of the animals may be disrupted by insertion or deletion mutation. After test substances have been administered to the animals, the growth of tumors must be assessed. If the test substance prevents or suppresses the growth of tumors, then the test substance is a candidate therapeutic agent for the treatment of FAP and/or sporadic cancers.

Polypeptides which have APC activity can be supplied to cells which carry mutant or missing APC alleles. The sequence of the APC protein is disclosed in Figure 3 or 7 (SEQ ID NO: 7 or 1). These two sequences differ slightly and appear to indicate the existence of two different forms of the APC protein. Protein can be produced by expression of the cDNA sequence in bacteria, for example, using known expression vectors. Alternatively, APC can be extracted from APC-producing mammalian cells such as brain cells. In addition, the techniques of synthetic chemistry can be employed to synthesize APC protein. Any of such techniques can provide the preparation of the present invention which comprises the APC protein. The preparation

is substantially free of other human proteins. This is most readily accomplished by synthesis in a microorganism or in vitro.

Active APC molecules can be introduced into cells by microinjection or by use of liposomes, for example. Alternatively, some such active molecules may be taken up by cells, actively or by diffusion. Extracellular application of APC gene product may be sufficient to affect tumor growth. Supply of molecules with APC activity should lead to a partial reversal of the neoplastic state. Other molecules with APC activity may also be used to effect such a reversal, for example peptides, drugs, or organic compounds.

The present invention also provides a preparation of antibodies immunoreactive with a human APC protein. The antibodies may be polyclonal or monoclonal and may be raised against native APC protein, APC fusion proteins, or mutant APC proteins. The antibodies should be immunoreactive with APC epitopes, preferably epitopes not present on other human proteins. In a preferred embodiment of the invention the antibodies will immunoprecipitate APC proteins from solution as well as react with APC protein on Western or immunoblots of polyacrylamide gels. In another preferred embodiment, the antibodies will detect APC proteins in paraffin or frozen tissue sections, using immunocytochemical techniques. Techniques for raising and purifying antibodies are well known in the art and any such techniques may be chosen to achieve the preparation of the invention.

Predisposition to cancers as in FAP and GS can be ascertained by testing any tissue of a human for mutations of the APC gene. For example, a person who has inherited a germline APC mutation would be prone to develop cancers. This can be determined by testing DNA from any tissue of the person's body. Most simply, blood can be drawn and DNA extracted from the cells of the blood. In addition, prenatal diagnosis can be accomplished by testing fetal cells, placental cells, or amniotic fluid for mutations of the APC gene. Alteration of a wild-type APC allele, whether for example, by point mutation or by deletion, can be detected by any of the means discussed above.

Molecules of cDNA according to the present invention are intron-free, APC gene coding molecules. They can be made by reverse

transcriptase using the APC mRNA as a template. These molecules can be propagated in vectors and cell lines as is known in the art. Such molecules have the sequence shown in SEQ ID NO: 7. The cDNA can also be made using the techniques of synthetic chemistry given the sequence disclosed herein.

A short region of homology has been identified between APC and the human m3 muscarinic acetylcholine receptor (mAChR). This homology was largely confined to 29 residues in which 6 out of 7 amino acids (EL(GorA)GLQA) were identical (See Figure 4). Initially, it was not known whether this homology is significant, because many other proteins had higher levels of global homology (though few had six out of seven contiguous amino acids in common). However, a study on the sequence elements controlling G protein activation by mAChR subtypes (Lechleiter et al., EMBO J., p. 4381 (1990)) has shown that a 21 amino acid region from the m3 mAChR completely mediated G protein specificity when substituted for the 21 amino acids of m2 mAChR at the analogous protein position. These 21 residues overlap the 19 amino acid homology between APC and m3 mAChR.

This connection between APC and the G protein activating region of mAChR is intriguing in light of previous investigations relating G proteins to cancer. For example, the RAS oncogenes, which are often mutated in colorectal cancers (Vogelstein, et al., N. Engl. J. Med., Vol. 319, p. 525 (1988); Bos et al., Nature Vol. 327, p. 293 (1987)), are members of the G protein family (Bourne, et al., Nature, Vol. 348, p. 125 (1990)) as is an *in vitro* transformation suppressor (Noda et al., Proc. Natl. Acad. Sci. USA, Vol. 86, p. 162 (1989)) and genes mutated in hormone producing tumors (Candis et al., Nature, Vol. 340, p. 692 (1989); Lyons et al., Science, Vol. 249, p. 655 (1990)). Additionally, the gene responsible for neurofibromatosis (presumably a tumor suppressor gene) has been shown to activate the GTPase activity of RAS (Xu et al., Cell, Vol. 63, p. 835 (1990); Martin et al., Cell, Vol. 63, p. 843 (1990); Ballester et al., Cell, Vol. 63, p. 851 (1990)). Another interesting link between G proteins and colon cancer involves the drug sulindac. This agent has been shown to inhibit the growth of benign colon tumors in patients with FAP, presumably by virtue of its activity as a

cyclooxygenase inhibitor (Waddell et al., J. Surg. Oncology 24(1), 83 (1983); Wadell, et al., Am. J. Surg., 157(1), 175 (1989); Charneau et al., Gastroenterologie Clinique et Biologique 14(2), 153 (1990)). Cyclooxygenase is required to convert arachidonic acid to prostaglandins and other biologically active molecules. G proteins are known to regulate phospholipase A2 activity, which generates arachidonic acid from phospholipids (Role et al., Proc. Natl. Acad. Sci. USA, Vol. 84, p. 3623 (1987); Kurachi et al., Nature, Vol. 337, 12 555 (1989)). Therefore we propose that wild-type APC protein functions by interacting with a G protein and involved in phospholipid metabolism.

The following are provided for exemplification purposes only and are not intended to limit the scope of the invention which has been described in broad terms above.

Example 1:

This example demonstrates the isolation of a 3.5 Mb region of human DNA linked to the FAP locus. Six genes are identified in this region, all of which are expressed in normal colon cells and in colorectal, lung, and bladder tumors.

The cosmid markers YN5.64 and YN5.48 have previously been shown to delimit an 8 cM region containing the locus for FAP (Nakamura et al., Am. J. Hum. Genet. Vol. 43, p. 638 (1988)). Further linkage and pulse-field gel electrophoresis (PFGE) analysis with additional markers has shown that the FAP locus is contained within a 4 cM region bordered by cosmids EF5.44 and L5.99. In order to isolate clones representing a significant portion of this locus, a yeast artificial chromosome (YAC) library was screened with various 5q21 markers. Twenty-one YAC clones, distributed within six contigs and including 5.5 Mb from the region between YN5.64 and YN5.48, were obtained (Figure 1A).

Three contigs encompassing approximately 4Mb were contained within the central portion of this region. The YAC's constituting these contigs, together with the markers used for their isolation and orientations, are shown in Figure 1. These YAC contigs were obtained in the following way. To initiate each contig, the sequence of a genomic

marker cloned from chromosome 5q21 was determined and used to design primers for PCR. PCR was then carried out on pools of YAC clones distributed in microtiter trays as previously described (Anand et al., *Nucleic Acids Research*, Vol. 18, p. 1951 (1990)). Individual YAC clones from the positive pools were identified by further PCR or hybridization based assays, and the YAC sizes were determined by PFGE.

To extend the areas covered by the original YAC clones, "chromosomal walking" was performed. For this purpose, YAC termini were isolated by a PCR based method and sequenced (Riley et al., *Nucleic Acids Research*, Vol. 18, p. 2887 (1990)). PCR primers based on these sequences were then used to rescreen the YAC library. For example, the sequence from an intron of the FER gene (Hao et al., *Mol. Cell. Biol.*, Vol. 9, p. 1587 (1989)) was used to design PCR primers for isolation of the 28EC1 and 5EH8 YACs. The termini of the 28EC1 YAC were sequenced to derive markers RHE28 and LHE28, respectively. The sequences of these two markers were then used to isolate YAC clones 15CH12 (from RHE28) and 40CF1 and 29EF1 (from LHE28). These five YAC's formed a contig encompassing 1200 kb (contig 1, Figure 1B).

Similarly, contig 2 was initiated using cosmid N5.66 sequences, and contig 3 was initiated using sequences both from the MCC gene and from cosmid EF5.44. A walk in the telomeric direction from YAC 14FH1 and a walk in the opposite direction from YAC 39GG3 allowed connection of the initial contig 3 clones through YAC 37HG4 (Figure 1B).

Multipoint linkage analysis with the various markers used to define the contigs, combined with PFGE analysis, showed that contigs 1 and 2 were centromeric to contig 3. These contigs were used as tools to orient and/or identify genes which might be responsible for FAP. Six genes were found to lie within this cluster of YAC's, as follows:

Contig #1: FER - The FER gene was discovered through its homology to the viral oncogene ABL (Hao et al., *supra*). It has an intrinsic tyrosine kinase activity, and in situ hybridization with an FER probe showed that the gene was located at 5q11-23 (Morris et al.,

Cytogenet. Cell. Genet., Vol. 53, p. 4, (1990)). Because of the potential role of this oncogene-related gene in neoplasia, we decided to evaluate it further with regards to the FAP locus. A human genomic clone from FER was isolated (MF 2.3) and used to define a restriction fragment length polymorphism (RFLP), and the RFLP in turn used to map FER by linkage analysis using a panel of three generation families. This showed that FER was very tightly linked to previously defined polymorphic markers for the FAP locus. The genetic mapping of FER was complemented by physical mapping using the YAC clones derived from FER sequences (Figure 1B). Analysis of YAC contig 1 showed that FER was within 600 kb of cosmid marker M5.28, which maps to within 1.5 Mb of cosmid L5.99 by PFGE of human genomic DNA. Thus, the YAC mapping results were consistent with the FER linkage data and PFGE analyses.

Contig 2: TB1 - TB1 was identified through a cross-hybridization approach. Exons of genes are often evolutionarily conserved while introns and intergenic regions are much less conserved. Thus, if a human probe cross-hybridizes strongly to the DNA from non-primate species, there is a reasonable chance that it contains exon sequences. Subclones of the cosmids shown in Figure 1 were used to screen Southern blots containing rodent DNA samples. A subclone of cosmid N5.66 (p 5.66-4) was shown to strongly hybridize to rodent DNA, and this clone was used to screen cDNA libraries derived from normal adult colon and fetal liver. The ends of the initial cDNA clones obtained in this screen were then used to extend the cDNA sequence. Eventually, 11 cDNA clones were isolated, covering 2314 bp. The gene detected by these clones was named TB1. Sequence analysis of the overlapping clones revealed an open reading frame (ORF) that extended for 1302 bp starting from the most 5' sequence data obtained (Figure 2A). If this entire open reading frame were translated, it would encode 434 amino acids. The product of this gene was not globally homologous to any other sequence in the current database but showed two significant local similarities to a family of ADP, ATP carrier/translocator proteins and mitochondrial brown fat uncoupling proteins which are widely distributed from yeast to mammals. These conserved regions of TB1

(underlined in Figure 2A) may define a predictive motif for this sequence family. In addition, TB1 appeared to contain a signal peptide (or mitochondrial targeting sequence) as well as at least 7 transmembrane domains.

Contig 3: MCC, TB2, SRP and APC - The MCC gene was also discovered through a cross-hybridization approach, as described previously (Kinzler et al., Science Vol. 251, p. 1366 (1991)). The MCC gene was considered a candidate for causing FAP by virtue of its tight genetic linkage to FAP susceptibility and its somatic mutation in sporadic colorectal carcinomas. However, mapping experiments suggested that the coding region of MCC was approximately 50 kb proximal to the centromeric end of a 200 kb deletion found in an FAP patient. MCC cDNA probes detected a 10 kb mRNA transcript on Northern blot analysis of which 4151 bp, including the entire open reading frame, have been cloned. Although the 3' non-translated portion or an alternatively spliced form of MCC might have extended into this deletion, it was possible that the deletion did not affect the MCC gene product. We therefore used MCC sequences to initiate a YAC contig, and subsequently used the YAC clones to identify genes 50 to 250 kb distal to MCC that might be contained within the deletion.

In a first approach, the insert from YAC24ED6 (Figure 1B) was radiolabelled and hybridized to a cDNA library from normal colon. One of the cDNA clones (YS39) identified in this manner detected a 3.1 kb mRNA transcript when used as a probe for Northern blot hybridization. Sequence analysis of the YS39 clone revealed that it encompassed 2283 nucleotides and contained an ORF that extended for 555 bp from the most 5' sequence data obtained. If all of this ORF were translated, it would encode 185 amino acids (Figure 2B). The gene detected by YS39 was named TB2. Searches of nucleotide and protein databases revealed that the TB2 gene was not identical to any previously reported sequences nor were there any striking similarities.

Another clone (YS11) identified through the YAC 24ED6 screen appeared to contain portions of two distinct genes. Sequences from one end of YS11 were identical to at least 180 bp of the signal recognition particle protein SRP19 (Lingelbach et al. Nucleic Acids Research,

Vol. 16, p. 9431 (1988). A second ORF, from the opposite end of clone YS11, proved to be identical to 78 bp of a novel gene which was independently identified through a second YAC-based approach. For the latter, DNA from yeast cells containing YAC 14FH1 (Figure 1B) was digested with EcoRI and subcloned into a plasmid vector. Plasmids that contained human DNA fragments were selected by colony hybridization using total human DNA as a probe. These clones were then used to search for cross-hybridizing sequences as described above for TB1, and the cross-hybridizing clones were subsequently used to screen cDNA libraries. One of the cDNA clones discovered in this way (FH38) contained a long ORF (2496 bp), 78 bp of which were identical to the above-noted sequences in YS11. The ends of the FH38 cDNA clone were then used to initiate cDNA walking to extend the sequence. Eventually, 85 cDNA clones were isolated from normal colon, brain and liver cDNA libraries and found to encompass 8973 nucleotides of contiguous transcript. The gene corresponding to this transcript was named APC. When used as probes for Northern blot analysis, APC cDNA clones hybridized to a single transcript of approximately 9.5 kb, suggesting that the great majority of the gene product was represented in the cDNA clones obtained. Sequences from the 5' end of the APC gene were found in YAC 37HG4 but not in YAC 14FH1. However, the 3' end of the APC gene was found in 14FH1 as well as 37HG4. The yeast artificial chromosome of the present invention designated YAC 37HG4 has been deposited with the National Collection of Industrial and Marine Bacteria (NCIMB), P.O. Box 31, 135 Abbey Road, Aberdeen AB9 8DG, Scotland, prior to the filing of this patent application. The NCIMB Accession Number of YAC clone YAC 37HG4 is 40353. Analogously, the 5' end of the MCC coding region was found in YAC clones 19AA9 and 26GC3 but not 24ED6 or 14FH1, while the 3' end displayed the opposite pattern. Thus, MCC and APC transcription units pointed in opposite directions, with the direction of transcription going from centromeric to telomeric in the case of MCC, and telomeric to centromeric in the case of APC. PFGE analysis of YAC DNA digested with various restriction endonucleases showed that TB2 and SRP were between MCC and APC, and that the 3' ends of the coding

regions of MCC and APC were separated by approximately 150 kb (Figure 1B).

Sequence analysis of the APC cDNA clones revealed an open reading frame of 8,535 nucleotides. The 5' end of the ORF contained a methionine codon (codon 1) that was preceded by an in-frame stop codon 9 bp upstream, and the 3' end was followed by several in-frame stop codons. The protein produced by initiation at codon 1 would contain 2,842 amino acids (Figure 3). The results of database searching with the APC gene product were quite complex due to the presence of large segments with locally biased amino acid compositions. In spite of this, APC could be roughly divided into two domains. The N-terminal 25% of the protein had a high content of leucine residues (12%) and showed local sequence similarities to myosins, various intermediate filament proteins (e.g., desmin, vimentin, neurofilaments) and *Drosophila* armadillo/human plakoglobin. The latter protein is a component of adhesive junctions (desmosomes) joining epithelial cells (Franke et al., Proc. Natl. Acad. Sci. U.S.A., Vol. 86, p. 4027 (1989); Perfer et al., Cell, Vol. 63, p. 1167 (1990)). The C-terminal 75% of APC (residues 731-2832) is 17% serine by composition with serine residues more or less uniformly distributed. This large domain also contains local concentrations of charged (mostly acidic) and proline residues. There was no indication of potential signal peptides, transmembrane regions, or nuclear targeting signals in APC suggesting a cytoplasmic localization.

To detect short similarities to APC, a database search was performed using the PAM-40 matrix (Altschul, J. Mol. Bio., Vol. 219, p. 555 (1991)). Potentially interesting matches to several proteins were found. The most suggestive of these involved the *ral2* gene product of yeast, which is implicated in the regulation of ras activity (Fukui et al., Mol. Cell. Biol., Vol. 9, p. 5617 (1989)). Little is known about how *ral2* might interact with ras but it is interesting to note the positively-charged character of this region in the context of the negatively-charged GAP interaction region of ras. A specific electrostatic interaction between ras and GAP-related proteins has been proposed.

Because of the proximity of the MCC and APC genes, and the fact that both are implicated in colorectal tumorigenesis, we searched for similarities between the two predicted proteins. Bourne has previously noted that MCC has the potential to form alpha helical coiled coils (Nature, Vol. 351, p. 188 (1991)). Lupas and colleagues have recently developed a program for predicting coiled coil potential from primary sequence data (Science, Vol. 252, p. 1162 (1991)) and we have used their program to analyze both MCC and APC. Analysis of MCC indicated a discontinuous pattern of coiled-coil domains separated by putative "hinge" or "spacer" regions similar to those seen in laminin and other intermediate filament proteins. Analysis of the APC sequence revealed two regions in the N-terminal domain which had strong coiled coil-forming potential, and these regions corresponded to those that showed local similarities with myosin and IF proteins on database searching. In addition, one other putative coiled coil region was identified in the central region of APC. The potential for both APC and MCC to form coiled coils is interesting in that such structures often mediate homo- and hetero-oligomerization.

Finally, it had previously been noted that MCC shared a short similarity with the region of the m3 muscarinic acetylcholine receptor (mAChR) known to regulate specificity of G-protein coupling. The APC gene also contained a local similarity to the region of the m3 mAChR that overlapped with the MCC similarity (Figure 4B). Although the similarities to ral2 (Figure 4A) and m3 mAChR (Figure 4B) were not statistically significant, they were intriguing in light of previous observations relating G-proteins to neoplasia.

Each of the six genes described above was expressed in normal colon mucosa, as indicated by their representation in colon cDNA libraries. To study expression of the genes in neoplastic colorectal epithelium, we employed reverse transcription-polymerase chain reaction (PCR) assays. Primers based on the sequences of FER, TB1, TB2, MCC, and APC were each used to design primers for PCR performed with cDNA templates. Each of these genes was found to be expressed in normal colon, in each of ten cell lines derived from colorectal cancers, and in tumor cell lines derived from lung and bladder tumors. The

ten colorectal cancer cell lines included eight from patients with sporadic CRC and two from patients with FAP.

Example 2

This example demonstrates a genetic analysis of the role of the FER gene in FAP and sporadic colorectal cancers.

We considered FER as a candidate because of its proximity to the FAP locus as judged by physical and genetic criteria (see Example 1), and its homology to known tyrosine kinases with oncogenic potential. Primers were designed to PCR-amplify the complete coding sequence of FER from the RNA of two colorectal cancer cell lines derived from FAP patients. cDNA was generated from RNA and used as a template for PCR. The primers used were 5'-AGAAGGATCCCTTGTGCAGTGTGGA-3' and 5'-GACAGGATCCTGAAGCTGAGTTTG-3'. The underlined nucleotides were altered from the true FER sequence to create BamHI sites. The cell lines used were JW and Difi, both derived from colorectal cancers of FAP patients. (C. Paraskeva, B.G. Buckle, D. Sheer, C.B. Wigley, *Int. J. Cancer* 34, 49 (1984); M.E. Gross et al., *Cancer Res.* 51, 1452 (1991). The resultant 2554 basepair fragments were cloned and sequenced in their entirety. The PCR products were cloned in the BamHI site of Bluescript SK (Stratagene) and pools of at least 50 clones were sequenced en masse using T7 polymerase, as described in Nigro et al., *Nature* 342, 705 (1989).

Only a single conservative amino acid change (GTG->CTG, creating a val to leu substitution at codon 439) was observed. The region surrounding this codon was then amplified from the DNA of individuals without FAP and this substitution was found to be a common polymorphism, not specifically associated with FAP. Based on these results, we considered it unlikely (though still possible) the FER gene was responsible for FAP. To amplify the regions surrounding codon 439, the following primers were used: 5'-TCAGAAAGTGCTGAAGAG-3' and 5'-GGAATAATTAGGTCTCCAA-3'. PCR products were digested with PstI, which yields a 50 bp fragment if codon 439 is leucine, but 26 and 24 bp fragments if it is valine. The primers used for sequencing were chosen from the FER cDNA sequence in Hao et al., supra.

Example 3

This example demonstrates the genetic analysis of MCC, TB2, SRP and APC in FAP and sporadic colorectal tumors. Each of these genes is linked and encompassed by contig 3 (see Figure 1).

Several lines of evidence suggested that this contig was of particular interest. First, at least three of the four genes in this contig were within the deleted region identified in two FAP patients. (See Example 5 infra.) Second, allelic deletions of chromosome 5q21 in sporadic cancers appeared to be centered in this region. (Ashton-Rickardt et al., *Oncogene*, in press; and Mild et al., *Jpn. J. Cancer Res.*, in press.) Some tumors exhibited loss of proximal RFLP markers (up to and potentially including the 5' end of MCC), but no loss of markers distal to MCC. Other tumors exhibited loss of markers distal to and perhaps including the 3' end of MCC, but no loss of sequences proximal to MCC. This suggested either that different ends of MCC were affected by loss in all such cases, or alternatively, that two genes (one proximal to and perhaps including MCC, the other distal to MCC) were separate targets of deletion. Third, clones from each of the six FAP region genes were used as probes on Southern blots containing tumor DNA from patients with sporadic CRC. Only two examples of somatic changes were observed in over 200 tumors studied: a rearrangement/deletion whose centromeric end was located within the MCC gene (Kinzler et al., supra) and an 800 bp insertion within the APC gene between nucleotides 4424 and 5584. Fourth, point mutations of MCC were observed in two tumors (Kinzler et al.) supra strongly suggesting that MCC was a target of mutation in at least some sporadic colorectal cancers.

Based on these results, we attempted to search for subtle alterations of contig 3 genes in patients with FAP. We chose to examine MCC and APC, rather than TB2 or SRP, because of the somatic mutations in MCC and APC noted above. To facilitate the identification of subtle alterations, the genomic sequences of MCC and APC exons were determined (see Table I). These sequences were used to design primers for PCR analysis of constitutional DNA from FAP patients.

We first amplified eight exons and surrounding introns of the MCC gene in affected individuals from 90 different FAP kindreds. The PCR products were analyzed by a ribonuclease (RNase) protein assay. In brief, the PCR products were hybridized to in vitro transcribed RNA probes representing the normal genomic sequences. The hybrids were digested with RNase A, which can cleave at single base pair mismatches within DNA-RNA hybrids, and the cleavage products were visualized following denaturing gel electrophoresis. Two separate RNase protection analyses were performed for each exon, one with the sense and one with the antisense strand. Under these conditions, approximately 40% of all mismatches are detectable. Although some amino acid variants of MCC were observed in FAP patients, all such variants were found in a small percentage of normal individuals. These variants were thus unlikely to be responsible for the inheritance of FAP.

We next examined three exons of the APC gene. The three exons examined included those containing nt 822-930, 931-1309, and the first 300 nt of the most distal exon (nt 1956-2256). PCR and RNase protection analysis were performed as described in Kinzler et al. supra, using the primers underlined in Table I. The primers for nt 1956-2256 were 5'-GCAAATCCTAAGAGAGAACA-3' and 5'-GATGGCAAGCTTGAGCCAG-3'.

In 90 kindreds, the RNase protection method was used to screen for mutations and in an additional 13 kindreds, the PCR products were cloned and sequenced to search for mutations not detectable by RNase protection. PCR products were cloned into a Bluescript vector modified as described in T.A. Holton and M.W. Graham, *Nucleic Acids Res.* 19, 1156 (1991). A minimum of 100 clones were pooled and sequenced. Five variants were detected among the 103 kindreds analyzed. Cloning and subsequent DNA sequencing of the PCR product of patient P21 indicated a C to T transition in codon 413 that resulted in a change from arginine to cysteine. This amino acid variant was not observed in any of 200 DNA samples from individuals without FAP. Cloning and sequencing of the PCR product from patients P24 and P34, who demonstrated the same abnormal RNase protection pattern indicated that

both had a C to T transition at codon 301 that resulted in a change from arginine (CGA) to a stop codon (TGA). This change was not present in 200 individuals without FAP. As this point mutation resulted in the predicted loss of the recognition site for the enzyme Taq I, appropriate PCR products could be digested with Taq I to detect the mutation. This allowed us to determine that the stop codon co-segregated with disease phenotype in members of the family of P24. The inheritance of this change in affected members of the pedigree provides additional evidence for the importance of the mutation.

Cloning and sequencing of the PCR product from FAP patient P93 indicated a C to G transversion at codon 279, also resulting in a stop codon (change from TCA to TGA). This mutation was not present in 200 individuals without FAP. Finally, one additional mutation resulting in a serine (TCA) to stop codon (TGA) at codon 712 was detected in a single patient with FAP (patient P60).

The five germline mutations identified are summarized in Table IIA, as well as four others discussed in Example 9. In addition to these germline mutations, we identified several somatic mutations of MCC and APC in sporadic CRC's. Seventeen MCC exons were examined in 90 sporadic colorectal cancers by RNase protection analysis. In each case where an abnormal RNase protection pattern was observed, the corresponding PCR products were cloned and sequenced. This led to the identification of six point mutations (two described previously) (Kinzler et al., *supra*), each of which was not found in the germline of these patients (Table IIB). Four of the mutations resulted in amino acid substitutions and two resulted in the alteration of splice site consensus elements. Mutations at analogous splice site positions in other genes have been shown to alter RNA processing in vivo and in vitro.

Three exons of APC were also evaluated in sporadic tumors. Sixty tumors were screened by RNase protection, and an additional 98 tumors were evaluated by sequencing. The exons examined included nt 822-930, 931-1309, and 1406-1545 (Table I). A total of three mutations were identified, each of which proved to be somatic. Tumor T27 contained a somatic mutation of CGA (arginine) to TGA (stop codon) at codon 33. Tumor T135 contained a GT to GC change at a splice donor

site. Tumor T34 contained a 5 bp insertion (CAGCC between codons 288 and 289) resulting in a stop at codon 291 due to a frameshift.

We serendipitously discovered one additional somatic mutation in a colorectal cancer. During our attempt to define the sequences and splice patterns of the MCC and APC gene products in colorectal epithelial cells, we cloned cDNA from the colorectal cancer cell line SW480. The amino acid sequence of the MCC gene from SW480 was identical to that previously found in clones from human brain. The sequence of APC in SW480 cells, however, differed significantly, in that a transition at codon 1338 resulted in a change from glutamine (CAG) to a stop codon (TAG). To determine if this mutation was somatic, we recovered DNA from archival paraffin blocks of the original surgical specimen (T201) from which the tumor cell line was derived 28 years ago.

DNA was purified from paraffin sections as described in S.E. Goetz, S.R. Hamilton, and B. Vogelstein. *Biochem. Biophys. Res. Comm.* 130, 118 (1985). PCR was performed as described in reference 24, using the primers 5'-GTTCCAGCAGTGTCACAG-3' and 5'-GGGAGATTTCGCTCCTGA-3'. A PCR product containing codon 1338 was amplified from the archival DNA and used to show that the stop codon represented a somatic mutation present in the original primary tumor and in cell lines derived from the primary and metastatic tumor sites, but not from normal tissue of the patient.

The ten point mutations in the MCC and APC genes so far discovered in sporadic CRCs are summarized in Table IIB. Analysis of the number of mutant and wild-type PCR clones obtained from each of these tumors showed that in eight of the ten cases, the wild-type sequence was present in approximately equal proportions to the mutant. This was confirmed by RFLP analysis using flanking markers from chromosome 5q which demonstrated that only two of the ten tumors (T135 and T201) exhibited an allelic deletion on chromosome 5q. These results are consistent with previous observations showing that 20-40% of sporadic colorectal tumors had allelic deletions of chromosome 5q. Moreover, these data suggest that mutations of 5q21 genes

are not limited to those colorectal tumors which contain allelic deletions of this chromosome.

Example 4

This example characterizes small, nested deletions in DNA from two unrelated FAP patients.

DNA from 40 FAP patients was screened with cosmids that had been mapped into a region near the APC locus to identify small deletions or rearrangements. Two of these cosmids, L5.71 and L5.79, hybridized with a 1200 kb NotI fragment in DNAs from most of the FAP patients screened.

The DNA of one FAP patient, 3214, showed only a 940 kb NotI fragment instead of the expected 1200 kb fragment. DNA was analyzed from four other members of the patient's immediate family; the 940 kb fragment was present in her affected mother (4711), but not in the other, unaffected family members. The mother also carried a normal 1200 kb NotI fragment that was transmitted to her two unaffected offspring. These observations indicated that the mutant polyposis allele is on the same chromosome as the 940 kb NotI fragment. A simple interpretation is that APC patients 3214 and 4711 each carry a 260 kb deletion within the APC locus.

If a deletion were present, then other enzymes might also be expected to produce fragments with altered mobilities. Hybridization of L5.79 to NruI-digested DNAs from both affected members of the family revealed a novel NruI fragment of 1300 kb, in addition to the normal 1200 kb NruI fragment. Furthermore, MluI fragments in patients 3214 and 4711 also showed an increase in size consistent with the deletion of an MluI site. The two chromosome 5 homologs of patient 3214 were segregated in somatic cell hybrid lines; HHW1155 (deletion hybrid) carried the abnormal homolog and HHW1159 (normal hybrid) carried the normal homolog.

Because patient 3214 showed only a 940 kb NotI fragment, she had not inherited the 1200 kb fragment present in the unaffected father's DNA. This observation suggests that he must be heterozygous for, and have transmitted, either a deletion of the L5.79 probe region or a variant NotI fragment too large to resolve on the gel system. As

expected, the hybrid cell line HHW1159, which carries the paternal homolog, revealed no resolved NotI fragment when probed with L5.79. However, probing of HHW1159 DNA with L5.79 following digestion with other enzymes did reveal restriction fragments, demonstrating the presence of DNA homologous to the probe. The father is, therefore, interpreted as heterozygous for a polymorphism at the NotI site, with one chromosome 5 having a 1200 kb NotI fragment and the other having a fragment too large to resolve consistently on the gel. The latter was transmitted to patient 3214.

When double digests were used to order restriction sites within the 1200 kb NotI fragment, L5.71 and L5.79 were both found to lie on a 550 kb NotI-NruI fragment and, therefore, on the same side of an NruI site in the 1200 kb NotI fragment. To obtain genomic representation of sequences present over the entire 1200 kb NotI fragment, we constructed a library of small-fragment inserts enriched for sequences from this fragment. DNA from the somatic cell hybrid HHW141, which contains about 40% of chromosome 5, was digested with NotI and electrophoresed under pulsed-field gel (PFG) conditions; EcoRI fragments from the 1200 kb region of this gel were cloned into a phage vector. Probe Map30 was isolated from this library. In normal individuals probe Map30 hybridizes to the 1200 kb NotI fragment and to a 200 kb NruI fragment. This latter hybridization places Map30 distal, with respect to the locations of L5.71 and L5.79, to the NruI site of the 550 kb NotI-NruI fragment.

Because Map30 hybridized to the abnormal, 1300 kb NruI fragment of patient 3214, the locus defined by Map30 lies outside the hypothesized deletion. Furthermore, in normal chromosomes Map30 identified a 200 kb NruI fragment and L5.79 identified a 1200 kb NruI fragment; the hypothesized deletion must, therefore, be removing an NruI site, or sites, lying between Map30 and L5.79, and these two probes must flank the hypothesized deletion. A restriction map of the genomic region, showing placement of these probes, is shown in Figure 5.

A NotI digest of DNA from another FAP patient, 3824, was probed with L5.79. In addition to the 1200 kb normal NotI fragment, a

fragment of approximately 1100 kb was observed, consistent with the presence of a 100 kb deletion in one chromosome 5. In this case, however, digestion with *Nru*I and *Mlu*I did not reveal abnormal bands, indicating that if a deletion were present, its boundaries must lie distal to the *Nru*I and *Mlu*I sites of the fragments identified by L5.79. Consistent with this expectation, hybridization of Map30 to DNA from patient 3824 identified a 760 kb *Mlu*I fragment in addition to the expected 860 kb fragment, supporting the interpretation of a 100 kb deletion in this patient. The two chromosome 5 homologs of patient 3824 were segregated in somatic cell hybrid lines; HHW1291 was found to carry only the abnormal homolog and HHW1290 only the normal homolog.

That the 860 kb *Mlu*I fragment identified by Map30 is distinct from the 830 kb *Mlu*I fragment identified previously by L5.79 was demonstrated by hybridization of Map30 and L5.79 to a *Not*I-*Mlu*I double digest of DNA from the hybrid cell (HHW1159) containing the nondeleted chromosome 5 homolog of patient 3214. As previously indicated, this hybrid is interpreted as missing one of the *Not*I sites that define the 1200 kb fragment. A 620 kb *Not*I-*Mlu*I fragment was seen with probe L5.79, and an 860 kb fragment was seen with Map30. Therefore, the 830 kb *Mlu*I fragment recognized by probe L5.79 must contain a *Not*I site in HHW1159 DNA; because the 860 kb *Mlu*I fragment remains intact, it does not carry this *Not*I site and must be distinct from the 830 kb *Mlu*I fragment.

Example 5

This example demonstrates the isolation of human sequences which span the region deleted in the two unrelated FAP patients characterized in Example 4.

A strong prediction of the hypothesis that patients 3214 and 3824 carry deletions is that some sequences present on normal chromosome 5 homologs would be missing from the hypothesized deletion homologs. Therefore, to develop genomic probes that might confirm the deletions, as well as to identify genes from the region, YAC clones from a contig seeded by cosmid L5.79 were localized from a library containing seven haploid human genome equivalents (Albertsen et al.,

Proc. Natl. Acad. Sci. U.S.A., Vol. 87, pp. 4256-4260 (1990)) with respect to the hypothesized deletions. Three clones, YACs 57B8, 310D8, and 183H12, were found to overlap the deleted region.

Importantly, one end of YAC 57B8 (clone AT57) was found to lie within the patient 3214 deletion. Inverse polymerase chain reaction (PCR) defined the end sequences of the insert of YAC 57B8. PCR primers based on one of these end sequences repeatedly failed to amplify DNA from the somatic cell hybrid (HHW1155) carrying the deleted homolog of patient 3214, but did amplify a product of the expected size from the somatic cell hybrid (HHW1159) carrying the normal chromosome 5 homolog. This result supported the interpretation that the abnormal restriction fragments found in the DNA of patient 3214 result from a deletion.

Additional support for the hypothesis of deletion in DNA from patient 3214 came from subcloned fragments of YAC 183H12, which spans the region in question. Y11, an EcoRI fragment cloned from YAC 183H12, hybridized to the normal, 1200 kb NotI fragment of patient 4711, but failed to hybridize to the abnormal, 940 kb NotI fragment of 4711 or to DNA from deletion cell line HHW1155. This result confirmed the deletion in patient 3214.

Two additional EcoRI fragments from YAC 183H12, Y10 and Y14, were localized within the patient 3214 deletion by their failure to hybridize to DNA from HHW1155. Probe Y10 hybridizes to a 150 kb NruI fragment in normal chromosome 5 homologs. Because the 3214 deletion creates the 1300 kb NruI fragment seen with the probes L5.79 and Map30 that flank the deletion, these NruI sites and the 150 kb NruI fragment lying between must be deleted in patient 3214. Furthermore, probe Y10 hybridizes to the same 620 kb NotI-MluI fragment seen with probe L5.79 in normal DNA, indicating its location as L5.79-proximal to the deleted MluI site and placing it between the MluI site and the L5.79-proximal NruI site. The MluI site must, therefore, lie between the NruI sites that define the 150 kb NruI fragment (see Figure 5).

Probe Y11 also hybridized to the 150 kb NruI fragment in the normal chromosome 5 homolog, but failed to hybridize to the 620 kb NotI-MluI fragment, placing it L5.79-distal to the MluI site, but

proximal to the second *Nru*I site. Hybridization to the same (860 kb) *Mlu*I fragment as Map30 confirmed the localization of probe Y11 L5.79-distal to the *Mlu*I site.

Probe Y14 was shown to be L5.79-distal to both deleted *Nru*I sites by virtue of its hybridization to the same 200 kb *Nru*I fragment of the normal chromosome 5 seen with Map30. Therefore, the order of these *Eco*RI fragments derived from YAC 183H12 and deleted in patient 3214, with respect to L5.79 and Map30, is L5.79-Y10-Y11-Y14-Map30.

The 100 kb deletion of patient 3824 was confirmed by the failure of aberrant restriction fragments in this DNA to hybridize with probe Y11, combined with positive hybridizations to probes Y10 and/or Y14. Y10 and Y14 each hybridized to the 1100 kb *Not*I fragment of patient 3824 as well as to the normal 1200 kb *Not*I fragment, but Y11 hybridized to the 1200 kb fragment only. In the *Mlu*I digest, probe Y14 hybridized to the 860 kb and 760 kb fragments of patient 3824 DNA, but probe Y11 hybridized only to the 860 kb fragment. We conclude that the basis for the alteration in fragment size in DNA from patient 3824 is, indeed, a deletion. Furthermore, because probes Y10 and Y14 are missing from the deleted 3214 chromosome, but present on the deleted 3824 chromosome, and they have been shown to flank probe Y11, the deletion in patient 3824 must be nested within the patient 3214 deletion.

Probes Y10, Y11, Y14 and Map30 each hybridized to YAC 310D8, indicating that this YAC spanned the patient 3824 deletion and at a minimum, most of the 3214 deletion. The YAC characterizations, therefore, confirmed the presence of deletions in the patients and provided physical representation of the deleted region.

Example 6

This example demonstrates that the MCC coding sequence maps outside of the region deleted in the two FAP patients characterized in Example 4.

An intriguing FAP candidate gene, MCC, recently was ascertained with cosmid L5.71 and was shown to have undergone mutation in colon carcinomas (Kinzler et al., *supra*). It was therefore of interest to

map this gene with respect to the deletions in APC patients. Hybridization of MCC probes with an overlapping series of YAC clones extending in either direction from L5.71 showed that the 3' end of MCC must be oriented toward the region of the two APC deletions.

Therefore, two 3' cDNA clones from MCC were mapped with respect to the deletions: clone 1CI (bp 2378-4181) and clone 7 (bp 2890-3560). Clone 1CI contains sequences from the C-terminal end of the open reading frame, which stops at nucleotide 2708, as well as 3' untranslated sequence. Clone 7 contains sequence that is entirely 3' to the open reading frame. Importantly, the entire 3' untranslated sequence contained in the cDNA clones consists of a single 2.5 kb exon. These two clones were hybridized to DNAs from the YACs spanning the FAP region. Clone 7 fails to hybridize to YAC 310D8, although it does hybridize to YACs 183H12 and 37B8; the same result was obtained with the cDNA 1CI. Furthermore, these probes did show hybridization to DNAs from both hybrid cell lines (HWW1159 and HWW1155) and the lymphoblastoid cell line from patient 3214, confirming their locations outside the deleted region. Additional mapping experiments suggested that the 3' end of the MCC cDNA clone contig is likely to be located more than 45 kb from the deletion of patient 3214 and, therefore, more than 100 kb from the deletion of patient 3824.

Example 7

This example identifies three genes within the deleted region of chromosome 5 in the two unrelated FAP patients characterized in Example 4.

Genomic clones were used to screen cDNA libraries in three separate experiments. One screening was done with a phage clone derived from YAC 310D8 known to span the 260 kb deletion of patient 3214. A large-insert phage library was constructed from this YAC; screening with Y11 identified λ 205, which mapped within both deletions. When clone λ 205 was used to probe a random-, plus oligo(dT)-, primed fetal brain cDNA library (approximately 300,000 phage), six cDNA clones were isolated and each of them mapped entirely within both deletions. Sequence analysis of these six clones formed a single cDNA contig, but did not reveal an extended open reading frame. One

of the six cDNAs was used to isolate more cDNA clones, some of which crossed the L5.71-proximal breakpoint of the 3824 deletion, as indicated by hybridization to both chromosome of this patient. These clones also contained an open reading frame, indicating a transcriptional orientation proximal to distal with respect to L5.71. This gene was named DP1 (deleted in polyposis 1). This gene is identical to TB2 described above.

cDNA walks yielded a cDNA contig of 3.0-3.5 kb, and included two clones containing terminal poly(A) sequences. This size corresponds to the 3.5 kb band seen by Northern analysis. Sequencing of the first 3163 bp of the cDNA contig revealed an open reading frame extending from the first base to nucleotide 631, followed by a 2.5 kb 3' untranslated region. The sequence surrounding the methionine codon at base 77 conforms to the Kozak consensus of an initiation methionine (Kozak, 1984). Failed attempts to walk farther, coupled with the similarity of the lengths of isolated cDNA and mRNA, suggested that the NH₂-terminus of the DP1 protein had been reached. Hybridization to a combination of genomic and YAC DNAs cut with various enzymes indicated the genomic coverage of DP1 to be approximately 30 kb.

Two additional probes for the locus, YS-11 and YS-39, which had been ascertained by screening of a cDNA library with an independent YAC probe identified with MCC sequences adjacent to L5.71, were mapped into the deletion region. YS-39 was shown to be a cDNA identical in sequence to DP1. Partial characterization of YS-11 had shown that 200 bp of DNA sequence at one end was identical to sequence coding for the 19 kd protein of the ribosomal signal recognition particle, SRP19 (Lingelbach et al., *supra*). Hybridization experiments mapped YS-11 within both deletions. The sequence of this clone, however, was found to be complex. Although 454 bp of the 1032 bp sequence of YS-11 were identical to the GenBank entry for the SRP19 gene, another 578 bp appended 5' to the SRP19 sequence was found to consist of previously unreported sequence containing no extended open reading frames. This suggested that YS-11 was either a chimeric clone containing two independent inserts or a clone of an incompletely processed or aberrant message. If YS-11 were a conventional chimeric clone, the

Independent segments would not be expected to map to the same physical region. The segments resulting from anomalous processing of a continuous transcript, however, would map to a single chromosomal region.

Inverse PCR with primers specific to the two ends of YS-11, the SRP19 end and the unidentified region, verified that both sequences map within the YAC 310D8; therefore, YS-11 is most likely a clone of an immature or anomalous mRNA species. Subsequently, both ends were shown to lie with the deleted region of patient 3824, and YS-11 was used to screen for additional cDNA clones.

Of the 14 cDNA clones selected from the fetal brain library, one clone, V5, was of particular interest in that it contained an open reading frame throughout, although it included only a short identity to the first 78 5' bases of the YS-11 sequence. Following the 78 bp of identical sequence, the two cDNA sequences diverged at an AG. Furthermore, divergence from genomic sequence was also seen after these 78 bp, suggesting the presence of a splice junction, and supporting the view that YS-11 represents an irregular message.

Starting with V5, successive 5' and 3' walks were performed; the resulting cDNA contig consisted of more than 100 clones, which defined a new transcript, DP2. Clones walking in the 5' direction crossed the 3824 deletion breakpoint farthest from LS71; since its 3' end is closer to this cosmid than its 5' end, the transcriptional orientation of DP2 is opposite to that of MCC and DP1.

The third screening approach relied on hybridization with a 120 kb MluI fragment from YAC 57B8. This fragment hybridizes with probe Y11 and completely spans the 100 kb deletion in patient 3824. The fragment was purified on two preparative PFGs, labeled, and used to screen a fetal brain cDNA library. A number of cDNA clones previously identified in the development of the DP1 and DP2 contigs were reascertained. However, 19 new cDNA clones mapped into the patient 3824 deletion. Analysis indicated that these 19 formed a new contig, DP3, containing a large open reading frame.

A clone from the 5' end of this new cDNA contig hybridized to the same EcoRI fragment as the 3' end of DP2. Subsequently, the DP2

and DP3 contigs were connected by a single 5' walking step from DP3, to form the single contig DP2.5. The complete nucleotide sequence of DP2.5 is shown in Figure 9.

The consensus cDNA sequence of DP2.5 suggests that the entire coding sequence of DP2.5 has been obtained and is 8532 bp long. The most 5' ATG codon occurs two codons from an in-frame stop and conforms to the Kozak initiation consensus (Kozak, Nucl. Acids. Res., Vol. 12, p. 857-872 1984). The 3' open reading frame breaks down over the final 1.8 kb, giving multiple stops in all frames. A poly(A) sequence was found in one clone approximately 1 kb into the 3' untranslated region, associated with a polyadenylation signal 33 bp upstream (position 9530). The open reading frame is almost identical to that identified as APC above.

An alternatively spliced exon at nucleotide 934 of the DP2.5 transcript is of potential interest. It was first discovered by noting that two classes of cDNA had been isolated. The more abundant cDNA class contains a 303 bp exon not included in the other. The presence in vivo of the two transcripts was verified by an exon connection experiment. Primers flanking the alternatively spliced exon were used to amplify, by PCR, cDNA prepared from various adult tissues. Two PCR products that differed in size by approximately 300 bases were amplified from all the tissues tested; the larger product was always more abundant than the smaller.

Example 8

This example demonstrates the primers used to identify subtle mutations in DP1, SRP19, and DP2.5.

To obtain DNA sequence adjacent to the exons of the genes DP1, DP2.5, and SRP19, sequencing substrate was obtained by inverse PCR amplification of DNAs from two YACs, 310D8 and 183H12, that span the deletions. Ligation at low concentration cyclized the restriction enzyme-digested YAC DNAs. Oligonucleotides with sequencing tails, designed in inverse orientation at intervals along the cDNAs, primed PCR amplification from the cyclized templates. Comparison of these DNA sequences with the cDNA sequences placed exon boundaries at the divergence points. SRP19 and DP1 were each shown to have five

exons. DP2.5 consisted of 15 exons. The sequences of the oligonucleotides synthesized to provide PCR amplification primers for the exons of each of these genes are listed in Table III. With the exception of exons 1, 3, 4, 9, and 15 of DP2.5 (see below), the primer sequences were located in intron sequences flanking the exons. The 5' primer of exon 1 is complementary to the cDNA sequence, but extends just into the 5' Kozak consensus sequence for the initiator methionine, allowing a survey of the translated sequences. The 5' primer of exon 3 is actually in the 5' coding sequences of this exon, as three separate intronic primers simply would not amplify. The 5' primer of exon 4 just overlaps the 5' end of this exon, and we thus fail to survey the 19 most 5' bases of this exon. For exon 9, two overlapping primer sets were used, such that each had one end within the exon. For exon 15, the large 3' exon of DP2.5, overlapping primer pairs were placed along the length of the exon; each pair amplified a product of 250-400 bases.

Example 9

This example demonstrates the use of single stranded conformation polymorphism (SSCP) analysis as described by Orita et al. Proc. Natl. Acad. Sci. U.S.A., Vol. 86, pp. 2766-70 (1989) and Genomics, Vol. 5, pp. 874-879 (1989) as applied to DP1, SRP19 and DP2.5.

SSCP analysis identifies most single- or multiple-base changes in DNA fragments up to 400 bases in length. Sequence alterations are detected as shifts in electrophoretic mobility of single-stranded DNA on nondenaturing acrylamide gels; the two complementary strands of a DNA segment usually resolve as two SSCP conformers of distinct mobilities. However, if the sample is from an individual heterozygous for a base-pair variant within the amplified segment, often three or more bands are seen. In some cases, even the sample from a homozygous individual will show multiple bands. Base-pair-change variants are identified by differences in pattern among the DNAs of the sample set.

Exons of the candidate genes were amplified by PCR from the DNAs of 61 unrelated FAP patients and a control set of 12 normal individuals. The five exons from DP1 revealed no unique conformers in the FAP patients, although common conformers were observed with exons

2 and 3 in some individuals of both affected and control sets, indicating the presence of DNA sequence polymorphisms. Likewise, none of the five exons of SRP19 revealed unique conformers in DNA from FAP patients in the test panel.

Testing of exons 1 through 14 and primer sets A through N of exon 15 of the DP2.5 gene, however, revealed variant conformers specific to FAP patients in exons 7, 8, 10, 11, and 15. These variants were in the unrelated patients 3746, 3460, 3827, 3712, and 3751, respectively. The PCR-SSCP procedure was repeated for each of these exons in the five affected individuals and in an expanded set of 48 normal controls. The variant bands were reproducible in the FAP patients but were not observed in any of the control DNA samples. Additional variant conformers in exons 11 and 15 of the DP2.5 gene were seen; however, each of these was found in both the affected and control DNA sets. The five sets of conformers unique to the FAP patients were sequenced to determine the nucleotide changes responsible for their altered mobilities. The normal conformers from the host individuals were sequenced also. Bands were cut from the dried acrylamide gels, and the DNA was eluted. PCR amplification of these DNAs provided template for sequencing.

The sequences of the unique conformers from exons 7, 8, 10, and 11 of DP2.5 revealed dramatic mutations in the DP2.5 gene. The sequence of the new mutation creating the exon 7 conformer in patient 3746 was shown to contain a deletion of two adjacent nucleotides, at positions 730 and 731 in the cDNA sequence (Figure 7). The normal sequence at this splice junction is CAGGGTCA (intronic sequence underlined), with the intron-exon boundary between the two repetitions of AG. The mutant allele in this patient has the sequence CAGGTCA. Although this change is at the 5' splice site, comparison with known consensus sequences of splice functions would suggest that a functional splice junction is maintained. If this new splice junction were functional, the mutation would introduce a frameshift that creates a stop codon 15 nucleotides downstream. If the new splice junction were not functional, messenger processing would be significantly altered.

To confirm the 2-base deletion, the PCR product from FAP patient 3746 and a control DNA were electrophoresed on an acrylamide-urea denaturing gel, along with the products of a sequencing reaction. The sample from patient 3746 showed two bands differing in size by 2 nucleotides, with the larger band identical in mobility to the control sample: this result was independent confirmation that patient 3746 is heterozygous for a 2 bp deletion.

The unique conformer found in exon 8 of patient 3460 was found to carry a C-T transition, at position 904 in the cDNA sequence of DP2.5 (shown in Figure 7), which replaced the normal sequence of CGA with TGA. This point mutation, when read in frame, results in a stop codon replacing the normal arginine codon. This single-base change had occurred within the context of a CG dimer, a potential hot spot for mutation (Barker et al., 1984).

The conformer unique to FAP patient 3827 in exon 10 was found to contain a deletion of one nucleotide (1367, 1368, or 1369) when compared to the normal sequence found in the other bands on the SSCP gel. This deletion, occurring within a set of three Ts, changed the sequence from CTTCA to CTTCA; this 1 base frameshift creates a downstream stop within 30 bases. The PCR product amplified from this patient's DNA also was electrophoresed on an acrylamide-urea denaturing gel, along with the PCR product from a control DNA and products from a sequencing reaction. The patient's PCR product showed two bands differing by 1 bp in length, with the larger identical in mobility to the PCR product from the normal DNA; this result confirmed the presence of a 1 bp deletion in patient 3827.

Sequence analysis of the variant conformer of exon 11 from patient 3712 revealed the substitution of a T by a G at position 1500, changing the normal tyrosine codon to a stop codon.

The pair of conformers observed in exon 15 of the DP2.5 gene for FAP patient 3751 also was sequenced. These conformers were found to carry a nucleotide substitution of C to G at position 5253, the third base of a valine codon. No amino acid change resulted from this substitution, suggesting that this conformer reflects a genetically silent polymorphism.

The observation of distinct inactivating mutations in the DP2.5 gene in four unrelated patients strongly suggested that DP2.5 is the gene involved in FAP. These mutations are summarized in Table IIA.

Example 10

This example demonstrates that the mutations identified in the DP2.5 (APC) gene segregate with the FAP phenotype.

Patient 3746, described above as carrying an APC allele with a frameshift mutation, is an affected offspring of two normal parents. Colonoscopy revealed no polyps in either parent nor among the patient's three siblings.

DNA samples from both parents, from the patient's wife, and from their three children were examined. SSCP analysis of DNA from both of the patient's parents displayed the normal pattern of conformers for exon 7, as did DNA from the patient's wife and one of his offspring. The two other children, however, displayed the same new conformers as their affected father. Testing of the patient and his parents with highly polymorphic VNTR (variable number of tandem repeat) markers showed a 99.98% likelihood that they are his biological parents.

These observations confirmed that this novel conformer, known to reflect a 2 bp deletion mutation in the DP2.5 gene, appeared spontaneously with FAP in this pedigree and was transmitted to two of the children of the affected individual.

Example 11

This example demonstrates polymorphisms in the APC gene which appear to be unrelated to disease (FAP).

Sequencing of variant conformers found among controls as well as individuals with APC has revealed the following polymorphisms in the APC gene: first, in exon 11, at position 1458, a substitution of T to C creating an RsaI restriction site but no amino acid change; and second, in exon 15, at positions 5037 and 5271, substitutions of A to G and G to T, respectively, neither resulting in amino acid substitutions. These nucleotide polymorphisms in the APC gene sequence may be useful for diagnostic purposes.

Example 12

This example shows the structure of the APC gene.

The structure of the APC gene is schematically shown in Figure 8, with flanking intron sequences indicated.

The continuity of the very large (6.5 kb), most 3' exon in DP2.5 was shown in two ways. First, inverse PCR with primers spanning the entire length of this exon revealed no divergence of the cDNA sequence from the genomic sequence. Second, PCR amplification with converging primers placed at intervals along the exon generated products of the same size whether amplified from the originally isolated cDNA, cDNA from various tissues, or genomic template. Two forms of exon 9 were found in DP2.5: one is the complete exon; and the other, labeled exon 9A, is the result of a splice into the interior of the exon that deletes bases 934 to 1236 in the mRNA and removes 101 amino acids from the predicted protein (see Figure 7).

Example 13

This example demonstrates the mapping of the FAP deletions with respect to the APC exons.

Somatic cell hybrids carrying the segregated chromosomes 5 from the 100 kb (HHW1291) and 260 kb (HHW1155) deletion patients were used to determine the distribution of the APC genes exons across the deletions. DNAs from these cell lines were used as template, along with genomic DNA from a normal control, for PCR-based amplification of the APC exons.

PCR analysis of the hybrids from the 260 kb deletion of patient 3214 showed that all but one (exon 1) of the APC exons are removed by this deletion. PCR analysis of the somatic cell hybrid HHW1291, carrying the chromosome 5 homolog with the 100 kb deletion from patient 3824, revealed that exons 1 through 9 are present but exons 10 through 15 are missing. This result placed the deletion breakpoint either between exons 9 and 10 or within exon 10.

Example 14

This example demonstrates the expression of alternately spliced APC messenger in normal tissues and in cancer cell lines.

Tissues that express the APC gene were identified by PCR amplification of cDNA made to mRNA with primers located within adjacent APC exons. In addition, PCR primers that flank the alternatively spliced exon 9 were chosen so that the expression pattern of both splice forms could be assessed. All tissue types tested (brain, lung, aorta, spleen, heart, kidney, liver, stomach, placenta, and colonic mucosa) and cultured cell lines (lymphoblasts, HL60, and choriocarcinoma) expressed both splice forms of the APC gene. We note, however, that expression by lymphocytes normally residing in some tissues, including colon, prevents unequivocal assessment of expression. The large mRNA, containing the complete exon 9 rather than only exon 9A, appears to be the more abundant message.

Northern analysis of poly(A)-selected RNA from lymphoblasts revealed a single band of approximately 10 kb, consistent with the size of the sequenced cDNA.

Example 15

This example discusses structural features of the APC protein predicted from the sequence.

The cDNA consensus sequence of APC predicts that the longer, more abundant form of the message codes for a 2842 or 2844 amino acid peptide with a mass of 311.8 kd. This predicted APC peptide was compared with the current data bases of protein and DNA sequences using both IntelliGenetics and GCG software packages. No genes with a high degree of amino acid sequence similarity were found. Although many short (approximately 20 amino acid) regions of sequence similarity were uncovered, none was sufficiently strong to reveal which, if any, might represent functional homology. Interestingly, multiple similarities to myosins and keratins did appear. The APC gene also was scanned for sequence motifs of known function; although multiple glycosylation, phosphorylation, and myristoylation sites were seen, their significance is uncertain.

Analysis of the APC peptide sequence did identify features important in considering potential protein structure. Hydropathy plots (Kyte and Doolittle, J. Mol. Biol. Vol. 157, pp. 105-132 (1982)) indicate that the APC protein is notably hydrophilic. No hydrophobic domains

suggesting a signal peptide or a membrane-spanning domain were found. Analysis of the first 1000 residues indicates that α -helical rods may form (Cohen and Parry, Trends Biochem. Sci. Vol. 77, pp. 245-248 (1986); there is a scarcity of proline residues and, there are a number of regions containing heptad repeats (apolar-X-X-apolar-X-X-X). Interestingly, in exon 9A, the deleted form of exon 9, two heptad repeat regions are reconnected in the proper heptad repeat frame, deleting the intervening peptide region. After the first 1000 residues, the high proline content of the remainder of the peptide suggests a compact rather than a rod-like structure.

The most prominent feature of the second 1000 residues is a 20 amino acid repeat that is iterated seven times with semiregular spacing (Table 4). The intervening sequences between the seven repeat regions contained 114, 116, 151, 205, 107, and 58 amino acids, respectively. Finally, residues 2200-24000 contain a 200 amino acid basic domain.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: ALBERTSEN, HANS
ANAND, RAKESH
CARLSON, MARY
GRODER, JOANNA
HEDGE, PHILIP J.
JOSLYN, GEOFF
KIMLER, KENNETH
MARKHAM, ALEXANDER F.
NAKANURA, YUSUKE
THLIVERIS, ANDREW

(ii) TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
GENE IN COLORECTAL CANCER IN HUMANS

(iii) NUMBER OF SEQUENCES: 94

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Banner, Birch, McKie & Beckett
(B) STREET: 1001 G Street, NW
(C) CITY: Washington
(D) STATE: D.C.
(E) COUNTRY: USA
(F) ZIP: 20001-4598

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 07/741,940
(B) FILING DATE: 08-AUG-1991
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Kagan, Sarah A.
(B) REGISTRATION NUMBER: 32,141
(C) REFERENCE/DOCKET NUMBER: 1107.035574

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 202-508-9100
(B) TELEFAX: 202-508-9295

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9606 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
(B) CLONE: DP2.5(APC)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 34..2562

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

GGACTCGGAA ATGAGCTCCA AGCGTAGCCA AGG ATG GCT GCA GCT TCA TAT GAT      54
           Met Ala Ala Ala Ser Tyr Asp
           1           5

CAG TTG TTA AAG CAA GTT CAG GCA CTC AAG ATG GAG AAC TCA AAT CTT      102
Gln Leu Leu Lys Gln Val Glu Ala Leu Lys Met Glu Asn Ser Asn Leu
           10           15           20

CGA CAA GAG CTA GAA GAT AAT TCC AAT CAT CTT ACA AAA CTC GAA ACT      150
Arg Gln Glu Leu Glu Asp Asn Ser Asn His Leu Thr Lys Leu Glu Thr
           25           30           35

GAG CCA TCT AAT ATG AAG GAA CTA CTT AAA CAA CTA CAA GGA AGT ATT      198
Glu Ala Ser Asn Met Lys Glu Val Leu Lys Gln Leu Gln Gly Ser Ile
           40           45           50           55

GAA GAT GAA GCT ATG GCT TCT TCT GGA CAG ATT GAT TTA TTA GAG CGT      246
Glu Asp Glu Ala Met Ala Ser Ser Gly Gln Ile Asp Leu Leu Glu Arg
           60           65           70

CTT AAA GAG CTT AAC TTA GAT AGC AGT AAT TTC CCT GGA GTA AAA CTG      294
Leu Lys Glu Leu Asn Leu Asp Ser Ser Asn Phe Pro Gly Val Lys Leu
           75           80           85

CGC TCA AAA ATG TCC CTC CCT TCT TAT GGA AGC CGG GAA GGA TCT GTA      342
Arg Ser Lys Met Ser Leu Arg Ser Tyr Gly Ser Arg Glu Gly Ser Val
           90           95           100

TCA AGC CGT TCT GCA GAG TGC AGT CCT GTT CCT ATG GCT TCA TTT CCA      390
Ser Ser Arg Ser Gly Glu Cys Ser Pro Val Pro Met Gly Ser Phe Pro
           105           110           115

AGA AGA CGC TTT GTA AAT GGA ACC AGA GAA AGT ACT GGA TAT TTA GAA      438
Arg Arg Gly Phe Val Asn Gly Ser Arg Glu Ser Thr Gly Tyr Leu Glu
           120           125           130           135

GAA CTT GAG AAA CAG AGC TCA TTG CTT CTT GCT GAT CTT CAC AAA GAA      486
Glu Leu Glu Lys Glu Arg Ser Leu Leu Leu Ala Asp Leu Asp Lys Glu
           140           145           150

GAA AAG GAA AAA GAC TGG TAT TAC GCT CAA CTT CAG AAT CTC ACT AAA      534
Glu Lys Glu Lys Asp Trp Tyr Tyr Ala Gln Leu Gln Asn Leu Thr Lys
           155           160           165

AGA ATA GAT AGT CTT CCT TTA ACT GAA AAT TTT TCC TTA CAA ACA GAT      582
Arg Ile Asp Ser Leu Pro Leu Thr Glu Asn Phe Ser Leu Gln Thr Asp
           170           175           180

TTG ACC AGA AGG CAA TTG GAA TAT GAA CCA AGC CAA ATC AGA GTT GCG      630
Leu Thr Arg Arg Gln Leu Glu Tyr Glu Ala Arg Gln Ile Arg Val Ala
           185           190           195

```

ATC GAA GAA CAA CTA GGT ACC TGC CAG GAT ATC CAA AAA CGA GCA CAG Met Glu Glu Gln Leu Gly Thr Cys Gln Asp Met Glu Lys Arg Ala Gln 200 205 210 215	678
CGA AGA ATA GCC ACA ATT CAG CAA ATC GAA AAG GAC ATA CTT CGT ATA Arg Arg Ile Ala Arg Ile Gln Gln Ile Glu Lys Asp Ile Leu Arg Ile 220 225 230	726
CGA CAG CTT TTA CAG TCC CAA GCA ACA GAA GCA CAG ACC TCA TCT CAG Arg Gln Leu Leu Gln Ser Gln Ala Thr Glu Ala Glu Arg Ser Ser Gln 235 240 245	774
AAC AAG CAT GAA ACC GCC TCA CAT GAT GCT CAG CCG CAG AAT GAA GGT Asn Lys His Glu Thr Gly Ser His Asp Ala Glu Arg Gln Asn Glu Gly 250 255 260	822
CAA CGA GTG GGA GAA ATC AAC ATG GCA ACT TCT GGT AAT GGT CAG GGT Gln Gly Val Gly Glu Ile Asn Met Ala Thr Ser Gly Asn Gly Gln Gly 265 270 275	870
TCA ACT ACA CGA ATC GAC CAT GAA ACA GCC AGT GTT TTG AGT TCT AGT Ser Thr Thr Arg Met Asp His Glu Thr Ala Ser Val Leu Ser Ser Ser 280 285 290 295	918
AGC ACA CAC TCT GCA CCT CGA AGC CTC ACA AGT CAT CTG GGA ACC AAG Ser Thr His Ser Ala Pro Arg Arg Leu Thr Ser His Leu Gly Thr Lys 300 305 310	966
GTG GAA ATG GTG TAT TCA TTG TTG TCA ATG CTT GGT ACT CAT GAT AAG Val Glu Met Val Tyr Ser Leu Leu Ser Met Leu Gly Thr His Asp Lys 315 320 325	1014
GAT GAT ATG TCG CGA ACT TTG CTA GGT ATG TCT AGC TCC CAA GAC AGC Asp Asp Met Ser Arg Thr Leu Leu Ala Met Ser Ser Ser Gln Asp Ser 330 335 340	1062
TGT ATA TCC ATG CGA CAG TCT GCA TGT CTT CCT CTC CTC ATC CAG CTT Cys Ile Ser Met Arg Gln Ser Gly Cys Leu Pro Leu Leu Ile Gln Leu 345 350 355	1110
TTA CAT GCC AAT GAC AAA GAC TCT GTA TTG TTG GGA AAT TCC CGC GCC Leu His Gly Asn Asp Lys Asp Ser Val Leu Leu Gly Asn Ser Arg Gly 360 365 370 375	1158
AGT AAA GAG GCT CGG GCC AGC GCC AGT GCA GCA CTC CAC AAC ATC ATT Ser Lys Glu Ala Arg Ala Arg Ala Ser Ala Ala Leu His Asn Ile Ile 380 385 390	1206
CAC TCA CAG CCT CAT GAC AAG AGA GGC AGC CGT GAA ATC CGA GTC CTT His Ser Gln Pro Asp Asp Lys Arg Gly Arg Arg Glu Ile Arg Val Leu 395 400 405	1254
CAT CTT TTG GAA CAG ATA CGC GCT TAC TGT GAA ACC TGT TGC CAG TGC His Leu Leu Glu Gln Ile Arg Ala Tyr Cys Glu Thr Cys Trp Glu Trp 410 415 420	1302
CAG GAA GCT CAT GAA CCA GCC ATG GAC CAG GAC AAA AAT CCA ATG CCA Gln Glu Ala His Glu Pro Gly Met Asp Gln Asp Lys Asn Pro Met Pro 425 430 435	1350

GCT CCT GTT GAA CAT CAG ATC TGT CCT CCT GTG TGT GTT CTA ATC AAA Ala Pro Val Glu His Gln Ile Cys Pro Ala Val Cys Val Leu Met Lys 440 445 450 455	1398
CTT TCA TTT GAT GAA GAG CAT AGA CAT GCA ATG AAT GAA CTA GGG GGA Leu Ser Phe Asp Glu Glu His Arg His Ala Met Asn Glu Leu Gly Gly 460 465 470	1446
CTA CAG GCC ATT GCA GAA TTA TTG CAA GTC CAC TGT CAA ATG TAT GCG Leu Gln Ala Ile Ala Glu Leu Leu Gln Val Asp Cys Glu Met Tyr Gly 475 480 485	1494
CTT ACT AAT GAC CAC TAC AGT ATT ACA CTA AGA CCA TAT GCT GGA ATG Leu Thr Asn Asp His Tyr Ser Ile Thr Leu Arg Arg Tyr Ala Gly Met 490 495 500	1542
GCT TTG ACA AAC TTC ACT TTT GGA GAT GTA GCC AAC AAG GCT ACG CTA Ala Leu Thr Asn Leu Thr Phe Gly Asp Val Ala Asn Lys Ala Thr Leu 505 510 515	1590
TCC TGT ATG AAA GGC TGC ATG AGA CCA CTT GTG GCC CAA CTA AAA TCT Cys Ser Met Lys Gly Cys Met Arg Ala Leu Val Ala Gln Leu Lys Ser 520 525 530 535	1638
GAA AGT GAA GAC TTA CAG CAG GTT ATT GCA AGT GTT TTG AGG AAT TTG Glu Ser Glu Asp Leu Gln Gln Val Ile Ala Ser Val Leu Arg Asn Leu 540 545 550	1686
TCT TGG CGA GCA GAT GTA AAT AGT AAA AAC ACC TTG CGA GAA CTT GGA Ser Trp Arg Ala Asp Val Asn Ser Lys Lys Thr Leu Arg Glu Val Gly 555 560 565	1734
ACT GTG AAA GCA TTG ATG GAA TGT GCT TTA GAA GTT AAA AAG GAA TCA Ser Val Lys Ala Leu Met Glu Cys Ala Leu Glu Val Lys Lys Glu Ser 570 575 580	1782
ACC CTC AAA AGC GTA TTG AGT GCC TTA TGG AAT TTG TCA GCA CAT TGC Thr Leu Lys Ser Val Leu Ser Ala Leu Trp Asn Leu Ser Ala His Cys 585 590 595	1830
ACT GAG AAT AAA GCT GAT ATA TGT GCT GTA GAT GGT GCA CTT GCA TTT Thr Glu Asn Lys Ala Asp Ile Cys Ala Val Asp Gly Ala Leu Ala Phe 600 605 610 615	1878
TTG GTT GGC ACT CTT ACT TAC CGG AGC CAG ACA AAC ACT TTA CCC ATT Leu Val Gly Thr Leu Thr Tyr Arg Ser Gln Thr Asn Thr Leu Ala Ile 620 625 630	1926
ATT GAA AGT GGA GGT GGG ATA TTA CGG AAT GTC TCC AGC TTG ATA GCT Ile Glu Ser Gly Gly Gly Ile Leu Arg Asn Val Ser Ser Leu Ile Ala 635 640 645	1974
ACA AAT CAG GAC CAC AGC CAA ATC CTA AGA GAG AAC AAC TGT CTA CAA Thr Asn Glu Asp His Arg Gln Ile Leu Arg Glu Asn Asn Cys Leu Gln 650 655 660	2022
ACT TTA TTA CAA CAC TTA AAA TCT CAT ACT TTG ACA ATA GTC ACT AAT Thr Leu Leu Gln His Leu Lys Ser His Ser Leu Thr Ile Val Ser Asn 665 670 675	2070

GCA TGT GGA ACT TTC TGG AAT CTC TCA GCA AGA AAT CCT AAA GAC CAG Ala Cys Gly Thr Leu Trp Asn Leu Ser Ala Arg Asn Pro Lys Asp Gln 680 685 690 695	2116
GAA GCA TTA TGG GAC ATG GCG GCA GTT AGC ATC CTC AAG AAC CTC ATT Glu Ala Leu Trp Asp Met Gly Ala Val Ser Met Leu Lys Asn Leu Ile 700 705 710	2166
CAT TCA AAG CAC AAA ATG ATT GCT ATG GGA AGT GCT GCA GCT TTA AGG His Ser Lys His Lys Met Ile Ala Met Gly Ser Ala Ala Ala Leu Arg 715 720 725	2214
AAT CTC ATG GCA AAT AGC CCT GCG AAG TAC AAG CAT GCC AAT ATT ATG Asn Leu Met Ala Asn Arg Pro Ala Lys Tyr Lys Asp Ala Asn Ile Met 730 735 740	2262
TCT CCT GGC TCA AGC TTC CCA TCT CTT CAT GTT AGC AAA CAA AAA GCC Ser Pro Gly Ser Ser Leu Pro Ser Leu His Val Arg Lys Gln Lys Ala 745 750 755	2310
CTA GAA GCA GAA TTA GAT GCT CAG CAC TTA TCA GAA ACT TTT GAC AAT Leu Glu Ala Glu Leu Asp Ala Gln His Leu Ser Glu Thr Phe Asp Asn 760 765 770 775	2358
ATA CAC AAT TTA AGT CCC AAG GCA TCT CAT CCT AGT AAG CAG AGA CAC Ile Asp Asn Leu Ser Pro Lys Ala Ser His Arg Ser Lys Gln Arg His 780 785 790	2406
AAG CAA AGT CTC TAT GGT GAT TAT GTT TTT GAC ACC AAT CGA CAT GAT Lys Gln Ser Leu Tyr Gly Asp Tyr Phe Asp Thr Asn Arg His Asp 795 800 805	2454
CAT AAT AGG TCA GAC AAT TTT AAT ACT GCG AAC ATG ACT GTC CTT TCA Asp Asn Arg Ser Asp Asn Phe Asn Thr Gly Asn Met Thr Val Leu Ser 810 815 820	2502
CCA TAT TTG AAT ACT ACA GTC TTA CCC AGC TCC TCT TCA TCA AGA GGA Pro Tyr Leu Asn Thr Thr Val Leu Pro Ser Ser Ser Ser Ser Arg Gly 825 830 835	2550
AGC TTA CAT AGT TCT CTT TCT GAA AAA GAT AGA AGT TTG GAG AGA GAA Ser Leu Asp Ser Ser Arg Ser Glu Lys Asp Arg Ser Leu Glu Arg Glu 840 845 850 855	2598
CGC CGA ATT GGT CTA GCG AAC TAC CAT CCA GCA ACA GAA AAT CCA GGA Arg Gly Ile Gly Leu Gly Asn Tyr His Pro Ala Thr Glu Asn Pro Gly 860 865 870	2646
ACT TCT TCA AAG CGA GGT TTC CAG ATC TCC ACC ACT GCA GCC CAG ATT Thr Ser Ser Lys Arg Gly Leu Gln Ile Ser Thr Thr Thr Ala Ala Gln Ile 875 880 885	2694
CCC AAA GTC ATG GAA GAA GTG TCA GCC ATT CAT ACC TCT CAG GAA GAC Ala Lys Val Met Glu Glu Val Ser Ala Ile His Thr Ser Gln Glu Asp 890 895 900	2742
AGA AGT TCT GCG TCT ACC ACT GAA TTA CAT TGT GTG ACA GAT GAG AGA Arg Ser Ser Gly Ser Thr Thr Glu Leu His Cys Val Thr Asp Glu Arg 905 910 915	2790

AAT CCA CTT AGA AGA AGC TCT GCT GCC CAT ACA CAT TCA AAC ACT TAC Asn Ala Leu Arg Arg Ser Ser Ala Ala His Thr His Ser Asn Thr Tyr 920 925 930	2838
AAT TTC ACT AAG TCG GAA AAT TCA AAT AGG ACA TGT TCT ATG CCT TAT Asn Phe Thr Lys Ser Glu Asn Ser Asn Arg Thr Cys Ser Met Pro Tyr 940 945 950	2886
GCC AAA TTA GAA TAC AAG AGA TCT TCA AAT GAT AGT TTA AAT AGT GTC Ala Lys Leu Glu Tyr Lys Arg Ser Ser Asn Asp Ser Leu Asn Ser Val 955 960 965	2934
AGT AGT AAT GAT GGT TAT GGT AAA AGA GGT CAA ATC AAA CCC TCG ATT Ser Ser Asn Asp Gly Tyr Gly Lys Arg Gly Gln Met Lys Pro Ser Ile 970 975 980	2982
GAA TCC TAT TCT CAA CAT GAT GAA AGT AAG TTT TGC AGT TAT GGT CAA Glu Ser Tyr Ser Glu Asp Asp Glu Ser Lys Phe Cys Ser Tyr Gly Gln 985 990 995	3030
TAC CCA CCC GAC CTA GCC CAT AAA ATA CAT AGT GCA AAT CAT ATC CAT Tyr Pro Ala Asp Leu Ala His Lys Ile His Ser Ala Asn His Met Asp 1000 1005 1010 1015	3078
GAT AAT GAT GGA GAA CTA GAT ACA CCA ATA AAT TAT AGT CTT AAA TAT Asp Asn Asp Gly Glu Leu Asp Thr Pro Ile Asn Tyr Ser Leu Lys Tyr 1020 1025 1030	3126
TCA GAT GAG CAG TTG AAC TCT GGA AGG CAA AGT CCT TCA CAG AAT GAA Ser Asp Glu Gln Leu Asn Ser Gly Arg Gln Ser Pro Ser Gln Asn Glu 1035 1040 1045	3174
AGA TGG GCA AGA CCC AAA CAC ATA ATA GAA GAT GAA ATA AAA CAA AGT Arg Trp Ala Arg Pro Lys His Ile Ile Glu Asp Glu Ile Lys Gln Ser 1050 1055 1060	3222
GAG CAA AGA CAA TCA AGG AAT CAA AGT ACA ACT TAT CCT GTT TAT ACT Glu Gln Arg Gln Ser Arg Asn Gln Ser Thr Thr Tyr Pro Val Tyr Thr 1065 1070 1075	3270
GAG AGC ACT GAT GAT AAA CAC CTC AAG TTC CAA CCA CAT TTT GGA CAG Glu Ser Thr Asp Asp Lys His Leu Lys Phe Gln Pro His Phe Gly Gln 1080 1085 1090 1095	3318
CAG GAA TGT GTT TCT CCA TAC AGG TCA CCG GGA GCC AAT GGT TCA GAA Gln Glu Cys Val Ser Pro Tyr Arg Ser Arg Gly Ala Asn Gly Ser Glu 1100 1105 1110	3366
ACA AAT CGA CTC GGT TCT AAT CAT GGA ATT AAT CAA AAT GTA AGC CAG Thr Asn Arg Val Gly Ser Asn His Gly Ile Asn Gln Asn Val Ser Gln 1115 1120 1125	3414
TCT TTG TGT CAA GAA GAT GAC TAT GAA GAT GAT AAG CCT ACC AAT TAT Ser Leu Cys Gln Glu Asp Asp Tyr Glu Asp Asp Lys Pro Thr Asn Tyr 1130 1135 1140	3462
AGT GAA CGT TAC TCT GAA GAA GAA CAG CAT GAA CAA CAA CAG AGA CCA Ser Glu Arg Tyr Ser Glu Glu Glu Gln His Glu Glu Glu Glu Arg Pro 1145 1150 1155	3510

ACA AAT TAT AGC ATA AAA TAT AAT GAA GAG AAA CCT CAT CTC GAT CAC Thr Asn Tyr Ser Ile Lys Tyr Asn Glu Glu Lys Arg His Val Asp Gln 1160 1165 1170 1175	3558
CCT ATT GAT TAT ACT TTA AAA TAT GCC ACA GAT ATT CCT TCA TCA CAG Pro Ile Asp Tyr Ser Leu Lys Tyr Ala Thr Asp Ile Pro Ser Ser Gln 1180 1185 1190	3606
AAA CAG TCA TTT TCA TTC TCA AAG AGT TCA TCT CCA CAA ACC AGT AAA Lys Gln Ser Phe Ser Phe Ser Lys Ser Ser Gly Gln Ser Ser Lys 1195 1200 1205	3654
ACC GAA CAT ATG TCT TCA ACC AGT GAG AAT ACC TCC ACA CCT TCA TCT Thr Glu His Met Ser Ser Ser Ser Glu Asn Thr Ser Thr Pro Ser Ser 1210 1215 1220	3702
AAT GCC AAG AGC CAG AAT CAG CTC CAT CCA AGT TCT CCA CAG AGT AGA Asn Ala Lys Arg Gln Asn Lys Leu His Pro Ser Ser Ala Gln Ser Arg 1225 1230 1235	3750
AGT GGT CAG CCT CAA AAG GCT GCC ACT TGC AAA GTT TCT TCT ATT AAC Ser Gly Gln Pro Gln Lys Ala Ala Thr Cys Lys Val Ser Ser Ile Asn 1240 1245 1250 1255	3798
CAA GAA ACA ATA CAG ACT TAT TGT GTA GAA GAT ACT CCA ATA TGT TTT Gln Glu Thr Ile Gln Thr Tyr Cys Val Glu Asp Thr Pro Ile Cys Phe 1260 1265 1270	3846
TCA AGA TGT AGT TCA TTA TCA TCT TTG TCA TCA GCT GAA GAT GAA ATA Ser Arg Cys Ser Ser Leu Ser Ser Leu Ser Ser Ala Glu Asp Glu Ile 1275 1280 1285	3894
CGA TGT AAT CAG ACG ACA CAG GAA GCA GAT TCT GCT AAT ACC CTC CAA Gly Cys Asn Gln Thr Thr Gln Glu Ala Asp Ser Ala Asn Thr Leu Gln 1290 1295 1300	3942
ATA CCA GAA ATA AAA GGA AAG ATT GGA ACT AGC TCA GCT GAA GAT CCT Ile Ala Glu Ile Lys Gly Lys Ile Gly Thr Arg Ser Ala Glu Asp Pro 1305 1310 1315	3990
GTG AGC GAA GTT CCA GCA GTG TCA CAG CAG GCT AGA ACC AAA TCC AGC Val Ser Glu Val Pro Ala Val Ser Gln His Pro Arg Thr Lys Ser Ser 1320 1325 1330 1335	4038
AGA CTC CAG GGT TCT AGT TTA TCT TCA GAA TCA GCC AGC CAC AAA GCT Arg Leu Gln Gly Ser Ser Leu Ser Ser Glu Ser Ala Arg His Lys Ala 1340 1345 1350	4086
GTT GAA TTT CCT TCA GGA GCG AAA TCT CCC TCC AAA AGT GGT GCT CAG Val Glu Phe Pro Ser Gly Ala Lys Ser Pro Ser Lys Ser Gly Ala Gln 1355 1360 1365	4134
ACA CCC AAA AGT CCA CCT GAA CAC TAT GTT CAG GAG ACC CCA CTC ATC Thr Pro Lys Ser Pro Pro Glu His Tyr Val Gln Glu Thr Pro Leu Met 1370 1375 1380	4182
TTT AGC AGA TGT ACT TCT GTG AGT TCA GTT GAT AGT TTT GAG AGT CGT Phe Ser Arg Cys Thr Ser Val Ser Ser Leu Asp Ser Phe Glu Ser Arg 1385 1390 1395	4230

TCC ATT CCC AGC TCC GTT CAG AGT GAA CCA TGC AGT CGA ATG GTA AGT Ser Ile Ala Ser Ser Val Gln Ser Glu Pro Cys Ser Gly Met Val Ser 1400 1405 1410 1415	4276
GGC ATT ATA AGC CCC AGT GAT CTT CCA GAT AGC CCT CGA CAA ACC ATG Gly Ile Ile Ser Pro Ser Asp Leu Pro Asp Ser Pro Gly Gln Thr Met 1420 1425 1430	4326
CCA CCA AGC AGA AGT AAA ACA CCT CCA CCA CCT CCT CAA ACA GCT CAA Pro Pro Ser Arg Ser Lys Thr Pro Pro Pro Pro Pro Gln Thr Ala Gln 1435 1440 1445	4374
ACC AAG CGA GAA GTA CCT AAA AAT AAA CCA CCT ACT CCT GAA AAG AGA Thr Lys Arg Glu Val Pro Lys Asn Lys Ala Pro Thr Ala Glu Lys Arg 1450 1455 1460	4422
GAG AGT CGA CCT AAG CAA GCT GCA GTA AAT GCT CCA CTT CAG AGC GTC Glu Ser Gly Pro Lys Gln Ala Val Asn Ala Val Gln Arg Val 1465 1470 1475	4470
CAG GTT CTT CCA GAT GCT GAT ACT TTA TTA CAT TTT CCC ACA GAA AGT Gln Val Leu Pro Asp Ala Asp Thr Leu Leu His Phe Ala Thr Glu Ser 1480 1485 1490 1495	4518
ACT CCA CAT GGA TTT TCT TGT TCA TCC ACC CTC AGT GCT CTC AGC CTC Thr Pro Asp Gly Phe Ser Cys Ser Ser Ser Leu Ser Ala Leu Ser Leu 1500 1505 1510	4566
GAT GAG CCA TTT ATA CAG AAA GAT GTC GAA TTA AGA ATA ATC CCT CCA Asp Glu Pro Phe Ile Gln Lys Asp Val Glu Leu Arg Ile Met Pro Pro 1515 1520 1525	4614
GTT CAG GAA AAT GAC AAT GGC AAT GAA ACA GAA TCA GAG CAG CCT AAA Val Gln Glu Asn Asp Asn Gly Asn Glu Thr Glu Ser Glu Gln Pro Lys 1530 1535 1540	4662
GAA TCA AAT GAA AAC CAA GAG AAA GAG GCA GAA AAA ACT ATT CAT TCT Glu Ser Asn Glu Asn Gln Glu Lys Glu Ala Glu Lys Thr Ile Asp Ser 1545 1550 1555	4710
GAA AAG GAC CTA TTA GAT GAT TCA CAT GAT GAT GAT ATT GAA ATA CTA Glu Lys Asp Leu Leu Asp Asp Ser Asp Asp Asp Ile Glu Ile Leu 1560 1565 1570 1575	4758
GAA GAA TGT ATT ATT TCT GCC ATG CCA ACA AAG TCA TCA CCT AAA GGC Glu Glu Cys Ile Ile Ser Ala Met Pro Thr Lys Ser Ser Arg Lys Gly 1580 1585 1590	4805
AAA AAG CCA GCC CAG ACT GCT TCA AAA TTA CCT CCA CCT GTC GCA AGC Lys Lys Pro Ala Gln Thr Ala Ser Lys Leu Pro Pro Pro Val Ala Arg 1595 1600 1605	4854
AAA CCA AGT CAG GTC CCT GTG TAC AAA CTT CTA CCA TCA CAA AAC AGC Lys Pro Ser Gln Leu Pro Val Tyr Lys Leu Leu Pro Ser Gln Asn Arg 1610 1615 1620	4902
TTG CAA CCC CAA AAG CAT GTT AGT TTT ACA CCG GCG GAT GAT ATG CCA Leu Gln Pro Gln Lys His Val Ser Phe Thr Pro Gly Asp Asp Met Pro 1625 1630 1635	4950

CCC GTC TAT TGT GTT CAA CGC ACA CCT ATA AAC TTT TCC ACA GCT ACA Arg Val Tyr Cys Val Glu Gly Thr Pro Ile Asn Phe Ser Thr Ala Thr 1640 1645 1650 1655	4996
TCT CTA AGT GAT CTA ACA ATC GAA TCC CCT CCA AAT GAG TTA GCT GCT Ser Leu Ser Asp Leu Thr Ile Glu Ser Pro Pro Asn Glu Leu Ala Ala 1660 1665 1670	5046
GGA CAA GGA GTT AGA GGA GGA GCA CAG TCA GGT GAA TTT GAA AAA CGA Gly Glu Gly Val Arg Gly Gly Ala Gln Ser Gly Glu Phe Glu Lys Arg 1675 1680 1685	5094
GAT ACC ATT CCT ACA GAA CGC AGA AGT ACA GAT GAG GCT CAA GGA GGA Asp Thr Ile Pro Thr Glu Gly Arg Ser Thr Asp Glu Ala Gln Gly Gly 1690 1695 1700	5142
AAA ACC TCA TCT GTA ACC ATA CCT GAA TTC GAT GAC AAT AAA GCA GAG Lys Thr Ser Ser Val Thr Ile Pro Glu Leu Asp Asp Asn Lys Ala Glu 1705 1710 1715	5190
GAA GGT GAT ATT CTT CCA GAA TGC ATT AAT TCT GCT ATC CCC AAA GGG Glu Gly Asp Ile Leu Ala Glu Cys Ile Asn Ser Ala Met Pro Lys Gly 1720 1725 1730 1735	5238
AAA AGT CAC AAG CCT TTC CGT GTC AAA AAG ATA ATG GAC CAG GTC CAG Lys Ser His Lys Pro Phe Arg Val Lys Lys Ile Met Asp Gln Val Gln 1740 1745 1750	5286
CAA GCA TCT GCG TCG TCT TCT CCA CCC AAC AAA AAT CAG TTA GAT GGT Gln Ala Ser Ala Ser Ser Ser Ala Pro Asn Lys Asn Gln Leu Asp Gly 1755 1760 1765	5334
AAG AAA AAG AAA CCA ACT TCA CCA GTA AAA CCT ATA CCA CAA AAT ACT Lys Lys Lys Lys Pro Thr Ser Pro Val Lys Pro Ile Pro Gln Asn Thr 1770 1775 1780	5382
GAA TAT AGC ACA CGT GTA AGA AAA AAT GCA GAC TCA AAG AAT AAT TTA Glu Tyr Arg Thr Arg Val Arg Lys Asn Ala Asp Ser Lys Asn Asn Leu 1785 1790 1795	5430
AAT GCT GAG ACA GTT TTC TCA GAC AAC AAA GAT TCA AAG AAA CAG AAT Asn Ala Glu Arg Val Phe Ser Asp Asn Lys Asp Ser Lys Lys Gln Asn 1800 1805 1810 1815	5478
TTC AAA AAT AAT TCC AAG CAC TTC AAT GAT AAG CTC CCA AAT AAT GAA Leu Lys Asn Asn Ser Lys Asp Phe Asn Asp Lys Leu Pro Asn Asn Glu 1820 1825 1830	5526
GAT AGA GTC AGA GGA AC TTT GCT TTT GAT TCA CCT CAT CAT TAC ACC Asp Arg Val Arg Gly Ser Phe Ala Phe Asp Ser Pro His His Tyr Thr 1835 1840 1845	5574
CCT ATT GAA GGA ACT CCT TAC TGT TTT TCA CGA AAT GAT TCT TTG AGT Pro Ile Glu Gly Thr Pro Tyr Cys Phe Ser Arg Asn Asp Ser Leu Ser 1850 1855 1860	5622
TCT CTA GAT TTT GAT GAT GAT GAT GTT GAC CTT TCC AGC GAA AAG GCT Ser Leu Asp Phe Asp Asp Asp Val Asp Leu Ser Arg Glu Lys Ala 1865 1870 1875	5670

CAA TTA AGA AAG GCA AAA GAA AAT AAG GAA TCA GAG GCT AAA GTT ACC Glu Leu Arg Lys Ala Lys Glu Asn Lys Glu Ser Glu Ala Lys Val Thr 1880 1885 1890 1895	5718
AGC CAC ACA GAA CTA ACC TCC AAC CAA CAA TCA GCT AAT AAG ACA CAA Ser His Thr Glu Leu Thr Ser Asn Gln Gln Ser Ala Asn Lys Thr Gln 1900 1905 1910	5766
GCT ATT GCA AAG CAG CCA ATA AAT CGA GGT CAG CCT AAA CCC ATA CTT Ala Ile Ala Lys Gln Pro Ile Asn Arg Gly Gln Pro Lys Pro Ile Leu 1915 1920 1925	5814
CAG AAA CAA TCC ACT TTT CCC CAG TCA TCC AAA GAC ATA CCA GAC AGA Gln Lys Gln Ser Thr Phe Pro Gln Ser Ser Lys Asp Ile Pro Asp Arg 1930 1935 1940	5862
GGC GCA GCA ACT GAT GAA AAG TTA CAG AAT TTT GCT ATT GAA AAT ACT Gly Ala Ala Thr Asp Glu Lys Leu Gln Asn Phe Ala Ile Glu Asn Thr 1945 1950 1955	5910
CCA GTT TGC TTT TCT CAT AAT TCC TCT CTC AGT TCT CTC ACT CAC ATT Pro Val Cys Phe Ser His Asn Ser Ser Leu Ser Ser Leu Ser Asp Ile 1960 1965 1970 1975	5958
GAC CAA GAA AAC AAC AAT AAA GAA AAT GAA CCT ATC AAA GAG ACT GAG Asp Gln Glu Asn Asn Asn Lys Glu Asn Glu Pro Ile Lys Glu Thr Glu 1980 1985 1990	6006
CCC CCT GAC TCA CAG GGA GAA CCA AGT AAA CCT CAA CCA TCA GGC TAT Pro Pro Asp Ser Gln Gly Glu Pro Ser Lys Pro Gln Ala Ser Gly Tyr 1995 2000 2005	6054
GCT CCT AAA TCA TTT CAT GTT GAA GAT ACC CCA GTT TGT TTC TCA AGA Ala Pro Lys Ser Phe His Val Glu Asp Thr Pro Val Cys Phe Ser Arg 2010 2015 2020	6102
AAC AGT TCT CTC AGT TCT CTT AGT ATT GAC TCT GAA GAT GAC CTC TTG Asn Ser Ser Leu Ser Ser Leu Ser Ile Asp Ser Glu Asp Asp Leu Leu 2025 2030 2035	6150
CAG GAA TGT ATA ACC TCC CCA ATC CCA AAA AAG AAA AAG CCT TCA AGA Gln Glu Cys Ile Ser Ser Ala Met Pro Lys Lys Lys Lys Pro Ser Arg 2040 2045 2050 2055	6198
CTC AAG GGT GAT AAT GAA AAA CAT AGT CCC AGA AAT ATG GGT GGC ATA Leu Lys Gly Asp Asn Glu Lys His Ser Pro Arg Asn Met Gly Gly Ile 2060 2065 2070	6246
TTA GGT GAA GAT CTC ACA CTT GAT TTG AAA GAT ATA CAG AGA CCA GAT Leu Gly Glu Asp Leu Thr Leu Asp Leu Lys Asp Ile Gln Arg Pro Asp 2075 2080 2085	6294
TCA GAA CAT GGT CTA TCC CCT GAT TCA GAA AAT TTT GAT TGG AAA GCT Ser Glu His Gly Leu Ser Pro Asp Ser Glu Asn Phe Asp Trp Lys Ala 2090 2095 2100	6342
ATT CAG GAA GGT GCA AAT TCC ATA CTA AGT AGT TTA CAT CAA GCT GCT Ile Gln Glu Gly Ala Asn Ser Ile Val Ser Ser Leu His Gln Ala Ala 2105 2110 2115	6390

GCT GCT GCA TGT TTA TCT AGA CAA GCT TCG TCT GAT TCA GAT TCC ATC Ala Ala Ala Cys Leu Ser Arg Gln Ala Ser Ser Asp Ser Asp Ser Ile 2120 2125 2130 2135	643E
GTT TCC CTG AAA TCA CGA ATC TCT CTG GGA TCA CCA TTT CAT CTT ACA Leu Ser Leu Lys Ser Gly Ile Ser Leu Gly Ser Pro Phe His Leu Thr 2140 2145 2150	648E
CCT GAT CAA GAA GAA AAA CCC TTT ACA AGT AAT AAA GGC CCA CGA ATT Pro Asp Gln Glu Glu Lys Pro Phe Thr Ser Asn Lys Gly Pro Arg Ile 2155 2160 2165	653E
CTA AAA CCA GCG GAG AAA AGT ACA TTT GAA AGT AAA AAG ATA GAA TCT Leu Lys Pro Gly Glu Lys Ser Thr Leu Glu Thr Lys Lys Ile Glu Ser 2170 2175 2180	658E
GAA AGT AAA GGA ATC AAA CGA CGA AAA AAA GTT TAT AAA AGT TTC ATT Glu Ser Lys Gly Ile Lys Gly Gly Lys Lys Val Tyr Lys Ser Leu Ile 2185 2190 2195	663E
ACT CGA AAA GTT CGA TCT AAT TCA GAA ATT TCA GGC CAA ATG AAA CAG Thr Gly Lys Val Arg Ser Asn Ser Glu Ile Ser Gly Gln Met Lys Gln 2200 2205 2210 2215	667E
CCC GTT CAA GCA AAC ATG CCT TCA ATC TCT CGA GGC AGC ACA ATG ATT Pro Leu Gln Ala Asn Met Pro Ser Ile Ser Arg Gly Arg Thr Met Ile 2220 2225 2230	672E
CAT ATT CCA GGA GTT CGA AAT AGC TCC TCA AOT ACA AGT CCT GTT TCT His Ile Pro Gly Val Arg Asn Ser Ser Ser Ser Thr Ser Pro Val Ser 2235 2240 2245	677E
AAA AAA GCG CCA CCC CTT AAG ACT CCA GCC TCC AAA AGC CCT AGT GAA Lys Lys Gly Pro Pro Leu Lys Thr Pro Ala Ser Lys Ser Pro Ser Glu 2250 2255 2260	682E
GGT CAA ACA GCC ACC ACT TCT CCT AGA GGA GCC AAG CCA TCT GTC AAA Gly Gln Thr Ala Thr Thr Ser Pro Arg Gly Ala Lys Pro Ser Val Lys 2265 2270 2275	687E
TCA GAA TTA AGC CCT GTT GCC AGC CAG ACA TCC CAA ATA GGT GGC TCA Ser Glu Leu Ser Pro Val Ala Arg Gln Thr Ser Gln Ile Gly Gly Ser 2280 2285 2290 2295	691E
AGT AAA GCA CCT TCT AGA TCA GGA TCT AGA GAT TCG ACC CCT TCA AGA Ser Lys Ala Pro Ser Arg Ser Gly Ser Arg Asp Ser Thr Pro Ser Arg 2300 2305 2310	696E
CCT GCC CAG CAA CCA TTA AGT AGA CCT ATA CAG TCT CCT GCC CGA AAC Pro Ala Gln Gln Pro Leu Ser Arg Pro Ile Gln Ser Pro Gly Arg Asn 2315 2320 2325	701E
TCA ATT TCC CCT GGT AGA AAT GGA ATA AGT CCT CCT AAC AAA TTA TCT Ser Ile Ser Pro Gly Arg Asn Gly Ile Ser Pro Pro Asn Lys Leu Ser 2330 2335 2340	706E
CAA CTT CCA AGC ACA TCA TCC CCT AGT ACT GCT TCA ACT AAG TCC TCA Gln Leu Pro Arg Thr Ser Ser Pro Ser Thr Ala Ser Thr Lys Ser Ser 2345 2350 2355	711E

GGT TCT GGA AAA ATC TCA TAT ACA TCT CCA GGT AGA CAG ATC AGC CAA Gly Ser Gly Lys Met Ser Tyr Thr Ser Pro Gly Arg Gln Met Ser Gln 2360 2365 2370 2375	7158
CAG AAC CTT ACC AAA CAA ACA GGT TTA TCC AAG AAT GCC AGT AGT ATT Gln Asn Leu Thr Lys Gln Thr Gly Leu Ser Lys Asn Ala Ser Ser Ile 2380 2385 2390	7206
CCA AGA AGT GAG TCT GCC TCC AAA GGA CTA AAT CAG ATC AAT AAT GGT Pro Arg Ser Glu Ser Ala Ser Lys Gly Leu Asn Gln Met Asn Asn Gly 2395 2400 2405	7254
AAT GGA GCC AAT AAA AAG GTA GAA CTT TCT AGA ATC TCT TCA ACT AAA Asn Gly Ala Asn Lys Lys Val Glu Leu Ser Arg Met Ser Ser Thr Lys 2410 2415 2420	7302
TCA AGT GGA AGT GAA TCT GAT AGA TCA GAA AGA CCT GTA TTA GTA CGC Ser Ser Gly Ser Glu Ser Asp Arg Ser Glu Arg Pro Val Leu Val Arg 2425 2430 2435	7350
CAG TCA AGT TTC ATC AAA GAA GCT CCA AGC CCA ACC TTA AGA AGA AAA Gln Ser Thr Phe Ile Lys Glu Ala Pro Ser Pro Thr Leu Arg Arg Lys 2440 2445 2450 2455	7398
TTC CAG GAA TCT GCT TCA TTT GAA TCT CTT TCT CCA TCA TCT AGA CCA Leu Glu Glu Ser Ala Ser Phe Glu Ser Leu Ser Pro Ser Ser Arg Pro 2460 2465 2470	7446
GCT TCT CCC ACT AGC TCC CAG GCA CAA ACT CCA GTT TTA AGT CCT TCC Ala Ser Pro Thr Arg Ser Gln Ala Gln Thr Pro Val Leu Ser Pro Ser 2475 2480 2485	7494
CTT CCT GAT ATC TCT CTA TCC ACA CAT TCG TCT GTT CAG GCT GGT GGA Leu Pro Asp Met Ser Leu Ser Thr His Ser Ser Val Gln Ala Gly Gly 2490 2495 2500	7542
TGG CGA AAA CTC CCA CCT AAT CTC AGT CCC ACT ATA GAG TAT AAT GAT Trp Arg Lys Leu Pro Pro Asn Leu Ser Pro Thr Ile Glu Tyr Asn Asp 2505 2510 2515	7590
GCA AGA CCA GCA AAG CGC CAT GAT ATT GCA CGC TCT CAT TCT GAA AGT Gly Arg Pro Ala Lys Arg His Asp Ile Ala Arg Ser His Ser Glu Ser 2520 2525 2530 2535	7638
CCT TCT AGA CTT CCA ATC AAT AGC TCA GGA ACC TGG AAA GGT GAG CAC Pro Ser Arg Leu Pro Ile Asn Arg Ser Gly Thr Trp Lys Arg Glu His 2540 2545 2550	7686
AGC AAA CAT TCA TCA TCC CTT CCT CGA GTA AGC ACT TGG AGA AGA ACT Ser Lys His Ser Ser Ser Leu Pro Arg Val Ser Thr Trp Arg Arg Thr 2555 2560 2565	7734
GCA AGT TCA TCT TCA ATT CTT TCT GCT TCA TCA GAA TCC AGT GAA AAA Gly Ser Ser Ser Ser Ile Leu Ser Ala Ser Ser Glu Ser Ser Glu Lys 2570 2575 2580	7782
GCA AAA AGT GAG GAT GAA AAA CAT GTC AAC TCT ATT TCA GGA ACC AAA Ala Lys Ser Glu Asp Glu Lys His Val Asn Ser Ile Ser Gly Thr Lys 2585 2590 2595	7830

CAA AGT AAA GAA AAC CAA GTA TCC CCA AAA GGA ACA TCC AGA AAA ATA Gln Ser Lys Glu Asn Gln Val Ser Ala Lys Gly Thr Trp Arg Lys Ile 2600 2605 2610 2615	7878
AAA GAA AAT GAA TTT TCT CCC ACA AAT AGT ACT TCT CAG ACC GTT TCC Lys Glu Asn Glu Phe Ser Pro Thr Asn Ser Thr Ser Gln Thr Val Ser 2620 2625 2630	7926
TCA GGT GCT ACA AAT GGT GCT GAA TCA AAG ACT CTA ATT TAT CAA ATC Ser Gly Ala Thr Asn Gly Ala Glu Ser Lys Thr Leu Ile Tyr Gln Met 2635 2640 2645	7974
CCA CCT GCT GTT TCT AAA ACA GAG GAT GTT TCG CTC AGA ATT GAG GAC Ala Pro Ala Val Ser Lys Thr Glu Asp Val Trp Val Arg Ile Glu Asp 2650 2655 2660	8022
TGT CCC ATT AAC AAT CCT AGA TCT GCA AGA TCT CCC ACA GGT AAT ACT Cys Pro Ile Asn Asn Pro Arg Ser Gly Arg Ser Pro Thr Gly Asn Thr 2665 2670 2675	8070
CCC CCG GTG ATT GAC AGT GTT TCA GAA AAG CCA AAT CCA AAC ATT AAA Pro Pro Val Ile Asp Ser Val Ser Glu Lys Ala Asn Pro Asn Ile Lys 2680 2685 2690 2695	8118
GAT TCA AAA GAT AAT CAG GCA AAA CAA AAT GTG GGT AAT GGC AGT GTT Asp Ser Lys Asp Asn Gln Ala Lys Gln Asn Val Gly Asn Gly Ser Val 2700 2705 2710	8166
CCC ATG CGT ACC GTG GGT TTC GAA AAT CCG CTC ACC TCC TTT ATT CAG Pro Met Arg Thr Val Gly Leu Glu Asn Arg Leu Thr Ser Phe Ile Gln 2715 2720 2725	8214
GTG GAT GCC CCT GAC CAA AAA GGA ACT GAG ATA AAA CCA GGA CAA AAT Val Asp Ala Pro Asp Gln Lys Gly Thr Glu Ile Lys Pro Gly Gln Asn 2730 2735 2740	8262
AAT CCT GTC CCT GTA TCA GAG ACT AAT GAA AGT CCT ATA GTG CAA CGT Asn Pro Val Pro Val Ser Glu Thr Asn Glu Ser Pro Ile Val Glu Arg 2745 2750 2755	8310
ACC CCA TTC AGT TCT AGC AGC TCA AGC AAA CAC AGT TCA CCT AGT GGG Thr Pro Phe Ser Ser Ser Ser Ser Lys His Ser Ser Pro Ser Gly 2760 2765 2770 2775	8358
ACT GTT GCT GCC AGA GTG ACT CCT TTT AAT TAC AAC CCA AGC CCT AGC Thr Val Ala Ala Arg Val Thr Pro Phe Asn Tyr Asn Pro Ser Pro Arg 2780 2785 2790	8406
AAA AGC AGC GCA GAT AGC ACT TCA GCT CCG CCA TCT CAG ATC CCA ACT Lys Ser Ser Ala Asp Ser Thr Ser Ala Arg Pro Ser Gln Ile Pro Thr 2795 2800 2805	8454
CCA GTG AAT AAC AAC ACA AAG AAG CCA GAT TCC AAA ACT GAC AGC ACA Pro Val Asn Asn Asn Thr Lys Lys Arg Asp Ser Lys Thr Asp Ser Thr 2810 2815 2820	8502
GAA TCC AGT GGA ACC CAA AGT CCT AAG CCG CAT TCT GCG TCT TAC CTT Glu Ser Ser Gly Thr Gln Ser Pro Lys Arg His Ser Gly Ser Tyr Leu 2825 2830 2835	8550

CTC ACA TCT GTT TAAAAGAGAG GAAGAATGAA ACTAAGAAAA TTCTATGTTA 8602
 Val Thr Ser Val
 2840
 ATTACAAC TG CTATATAGAC ATTTTCITTC AAATGAACT TTAAGACT GAAAAATTTT 8662
 GTAAATAGGT TTGATTCTTG TTAGAGGGTT TTTGTTCTGG AAGCCATATT TCATAGTATA 8722
 CTTTGTCTTC ACTGCTCTTA TTTEGGAGC CACTCTTGAT GGTTAGGAAA AAATAGAAAG 8782
 CCAAGTATGT TTGTACAGTA TGTTTTACAT GTATTTAAAG TAGCATCCCA TCCCAACTTC 8842
 CTTAATTATT GCTTGTCTAA AAATATGAAC ACTACAGATA GGAATATGA TATATTGCTG 8902
 TTATCAATCA TTTCTAGATT ATAACTGAC TAACTTACA TCAGCGGAAA ATTGGTATTT 8962
 ATGCAAAAA AAAATGTTTT TGTCTTGTG AGTCCATCTA ACATCATAAT TAATCATGTG 9022
 GCTGTGAAAT TCACAGTAAT ATGTTTCCCC ATGAACAAGT TTACCCAGCC TGCTTTGGTT 9082
 ACTCCATCAA TGAAC TGAT GGTTCATTT CAGAATAAT GATTACAGT TATGTGGTCA 9142
 CATGATGTGC ATAGAGATAG CTACAGTGA ATAAATTACA CTATTTTGTG CTCCAAACAA 9202
 AACAAAAATC TGTGTAAC TG TAAACATTG AATGAAACTA TTTTACCTGA ACTAGATTTT 9262
 ATCTGAAAGT AGGTAGAAAT TTGCTATGC TGTAAATTTGT TGTATATTCT GGTATTTGAG 9322
 CTCAGATGGC TGCTCTTTAT TAATGACACA TGAATTTGT CTCAACAGAA ACTAAATGAA 9382
 CATTTCAGAA TAAATTATTC CTGTATGTAA ACTGTTACTG AAATGGGTAT TTGTTTGAAG 9442
 GGTTTGTTTC ACATTGTAT TAATTAATTG TTTAAAATGC CTCTTTTAAA AGCTTATATA 9502
 AATTTTTTCT TCAGCTTCTA TGCATTAAGA GTAAAATTCC TCTTACTGTA ATAAAAACAT 9562
 TGAAGAAGAC TGTTCCTACT TAACCATTCC ATGCGTTGGC ACTT 9606

(2) INFORMATION FOR SEQ ID NO:2:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2843 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala Ala Ala Ser Tyr Asp Gln Leu Leu Lys Gln Val Glu Ala Leu
 1 5 10 15
 Lys Met Glu Asn Ser Asn Leu Arg Gln Glu Leu Glu Asp Asn Ser Asn
 20 25 30
 His Leu Thr Lys Leu Glu Thr Glu Ala Ser Asn Met Lys Glu Val Leu
 35 40 45
 Lys Gln Leu Gln Gly Ser Ile Glu Asp Glu Ala Met Ala Ser Ser Gly
 50 55 60

Gln Ile Asp Leu Leu Glu Arg Leu Lys Glu Leu Asn Leu Asp Ser Ser
 65 70 75 80
 Asn Phe Pro Gly Val Lys Leu Arg Ser Lys Met Ser Leu Arg Ser Tyr
 85 90 95
 Gly Ser Arg Glu Gly Ser Val Ser Ser Arg Ser Gly Glu Cys Ser Pro
 100 105 110
 Val Pro Met Gly Ser Phe Pro Arg Arg Gly Phe Val Asn Gly Ser Arg
 115 120 125
 Glu Ser Thr Gly Tyr Leu Glu Glu Leu Glu Lys Glu Arg Ser Leu Leu
 130 135 140
 Leu Ala Asp Leu Asp Lys Glu Glu Lys Glu Lys Asp Trp Tyr Tyr Ala
 145 150 155 160
 Gln Leu Gln Asn Leu Thr Lys Arg Ile Asp Ser Leu Pro Leu Thr Glu
 165 170 175
 Asn Phe Ser Leu Gln Thr Asp Leu Thr Arg Arg Gln Leu Glu Tyr Glu
 180 185 190
 Ala Arg Gln Ile Arg Val Ala Met Glu Glu Gln Leu Gly Thr Cys Gln
 195 200 205
 Asp Met Glu Lys Arg Ala Gln Arg Arg Ile Ala Arg Ile Gln Gln Ile
 210 215 220
 Glu Lys Asp Ile Leu Arg Ile Arg Gln Leu Leu Gln Ser Gln Ala Thr
 225 230 235 240
 Glu Ala Glu Arg Ser Ser Gln Asn Lys His Glu Thr Gly Ser His Asp
 245 250 255
 Ala Glu Arg Gln Asn Glu Gly Gln Gly Val Gly Glu Ile Asn Met Ala
 260 265 270
 Thr Ser Gly Asn Gly Gln Gly Ser Thr Thr Arg Met Asp His Glu Thr
 275 280 285
 Ala Ser Val Leu Ser Ser Ser Ser Thr His Ser Ala Pro Arg Arg Leu
 290 295 300
 Thr Ser His Leu Gly Thr Lys Val Glu Met Val Tyr Ser Leu Leu Ser
 305 310 315 320
 Met Leu Gly Thr His Asp Lys Asp Asp Met Ser Arg Thr Leu Leu Ala
 325 330 335
 Met Ser Ser Ser Gln Asp Ser Cys Ile Ser Met Arg Gln Ser Gly Cys
 340 345 350
 Leu Pro Leu Leu Ile Gln Leu Leu His Gly Asn Asp Lys Asp Ser Val
 355 360 365
 Leu Leu Gly Asn Ser Arg Gly Ser Lys Glu Ala Arg Ala Arg Ala Ser
 370 375 380
 Ala Ala Leu His Asn Ile Ile His Ser Gln Pro Asp Asp Lys Arg Gly
 385 390 395 400

Arg Arg Glu Ile Arg Val Leu His Leu Leu Glu Gln Ile Arg Ala Tyr
 405 410 415
 Cys Glu Thr Cys Trp Glu Trp Gln Glu Ala His Glu Pro Gly Met Asp
 420 425 430
 Gln Asp Lys Asn Pro Met Pro Ala Pro Val Glu His Gln Ile Cys Pro
 435 440 445
 Ala Val Cys Val Leu Met Lys Leu Ser Phe Asp Glu Glu His Arg His
 450 455 460
 Ala Met Asn Glu Leu Gly Gly Leu Gln Ala Ile Ala Glu Leu Leu Gln
 465 470 475 480
 Val Asp Cys Glu Met Tyr Gly Leu Thr Asn Asp His Tyr Ser Ile Thr
 485 490 495
 Leu Arg Arg Tyr Ala Gly Met Ala Leu Thr Asn Leu Thr Phe Gly Asp
 500 505 510
 Val Ala Asn Lys Ala Thr Leu Cys Ser Met Lys Gly Cys Met Arg Ala
 515 520 525
 Leu Val Ala Gln Leu Lys Ser Glu Ser Glu Asp Leu Gln Gln Val Ile
 530 535 540
 Ala Ser Val Leu Arg Asn Leu Ser Trp Arg Ala Asp Val Asn Ser Lys
 545 550 555 560
 Lys Thr Leu Arg Glu Val Gly Ser Val Lys Ala Leu Met Glu Cys Ala
 565 570 575
 Leu Glu Val Lys Lys Glu Ser Thr Leu Lys Ser Val Leu Ser Ala Leu
 580 585 590
 Trp Asn Leu Ser Ala His Cys Thr Glu Asn Lys Ala Asp Ile Cys Ala
 595 600 605
 Val Asp Gly Ala Leu Ala Phe Leu Val Gly Thr Leu Thr Tyr Arg Ser
 610 615 620
 Gln Thr Asn Thr Leu Ala Ile Ile Glu Ser Gly Gly Gly Ile Leu Arg
 625 630 635 640
 Asn Val Ser Ser Leu Ile Ala Thr Asn Glu Asp His Arg Gln Ile Leu
 645 650 655
 Arg Glu Asn Asn Cys Leu Gln Thr Leu Leu Gln His Leu Lys Ser His
 660 665 670
 Ser Leu Thr Ile Val Ser Asn Ala Cys Gly Thr Leu Trp Asn Leu Ser
 675 680 685
 Ala Arg Asn Pro Lys Asp Gln Glu Ala Leu Trp Asp Met Gly Ala Val
 690 695 700
 Ser Met Leu Lys Asn Leu Ile His Ser Lys His Lys Met Ile Ala Met
 705 710 715 720
 Gly Ser Ala Ala Ala Leu Arg Asn Leu Met Ala Asn Arg Pro Ala Lys
 725 730 735

Tyr Lys Asp Ala Asn Ile Met Ser Pro Gly Ser Ser Leu Pro Ser Leu
 740 745 750
 His Val Arg Lys Gln Lys Ala Leu Glu Ala Glu Leu Asp Ala Gln His
 755 760 765
 Leu Ser Glu Thr Phe Asp Asn Ile Asp Asn Leu Ser Pro Lys Ala Ser
 770 775 780
 His Arg Ser Lys Gln Arg His Lys Gln Ser Leu Tyr Gly Asp Tyr Val
 785 790 795 800
 Phe Asp Thr Asn Arg His Asp Asp Asn Arg Ser Asp Asn Phe Asn Thr
 805 810 815
 Gly Asn Met Thr Val Leu Ser Pro Tyr Leu Asn Thr Thr Val Leu Pro
 820 825 830
 Ser Ser Ser Ser Ser Arg Gly Ser Leu Asp Ser Ser Arg Ser Glu Lys
 835 840 845
 Asp Arg Ser Leu Glu Arg Glu Arg Gly Ile Gly Leu Gly Asn Tyr His
 850 855 860
 Pro Ala Thr Glu Asn Pro Gly Thr Ser Ser Lys Arg Gly Leu Gln Ile
 865 870 875 880
 Ser Thr Thr Ala Ala Gln Ile Ala Lys Val Met Glu Glu Val Ser Ala
 885 890 895
 Ile His Thr Ser Gln Glu Asp Arg Ser Ser Gly Ser Thr Thr Glu Leu
 900 905 910
 His Cys Val Thr Asp Glu Arg Asn Ala Leu Arg Arg Ser Ser Ala Ala
 915 920 925
 His Thr His Ser Asn Thr Tyr Asn Phe Thr Lys Ser Glu Asn Ser Asn
 930 935 940
 Arg Thr Cys Ser Met Pro Tyr Ala Lys Leu Glu Tyr Lys Arg Ser Ser
 945 950 955 960
 Asn Asp Ser Leu Asn Ser Val Ser Ser Asn Asp Gly Tyr Gly Lys Arg
 965 970 975
 Gly Gln Met Lys Pro Ser Ile Glu Ser Tyr Ser Glu Asp Asp Glu Ser
 980 985 990
 Lys Phe Cys Ser Tyr Gly Gln Tyr Pro Ala Asp Leu Ala His Lys Ile
 995 1000 1005
 His Ser Ala Asn His Met Asp Asp Asn Asp Gly Glu Leu Asp Thr Pro
 1010 1015 1020
 Ile Asn Tyr Ser Leu Lys Tyr Ser Asp Glu Gln Leu Asn Ser Gly Arg
 1025 1030 1035 1040
 Gln Ser Pro Ser Gln Asn Glu Arg Trp Ala Arg Pro Lys His Ile Ile
 1045 1050 1055
 Glu Asp Glu Ile Lys Gln Ser Glu Gln Arg Gln Ser Arg Asn Gln Ser
 1060 1065 1070

Thr Thr Tyr Pro Val Tyr Thr Glu Ser Thr Asp Asp Lys His Leu Lys
 1075 1080 1085
 Phe Gln Pro His Phe Gly Gln Gln Glu Cys Val Ser Pro Tyr Arg Ser
 1090 1095 1100
 Arg Gly Ala Asn Gly Ser Glu Thr Asn Arg Val Gly Ser Asn His Gly
 1105 1110 1115 1120
 Ile Asn Gln Asn Val Ser Gln Ser Leu Cys Gln Glu Asp Asp Tyr Glu
 1125 1130 1135
 Asp Asp Lys Pro Thr Asn Tyr Ser Glu Arg Tyr Ser Glu Glu Glu Gln
 1140 1145 1150
 His Glu Glu Glu Glu Arg Pro Thr Asn Tyr Ser Ile Lys Tyr Asn Glu
 1155 1160 1165
 Glu Lys Arg His Val Asp Gln Pro Ile Asp Tyr Ser Leu Lys Tyr Ala
 1170 1175 1180
 Thr Asp Ile Pro Ser Ser Gln Lys Gln Ser Phe Ser Phe Ser Lys Ser
 1185 1190 1195 1200
 Ser Ser Gly Gln Ser Ser Lys Thr Glu His Met Ser Ser Ser Ser Glu
 1205 1210 1215
 Asn Thr Ser Thr Pro Ser Ser Asn Ala Lys Arg Gln Asn Gln Leu His
 1220 1225 1230
 Pro Ser Ser Ala Gln Ser Arg Ser Gly Gln Pro Gln Lys Ala Ala Thr
 1235 1240 1245
 Cys Lys Val Ser Ser Ile Asn Gln Glu Thr Ile Gln Thr Tyr Cys Val
 1250 1255 1260
 Glu Asp Thr Pro Ile Cys Phe Ser Arg Cys Ser Ser Leu Ser Ser Leu
 1265 1270 1275 1280
 Ser Ser Ala Glu Asp Glu Ile Gly Cys Asn Gln Thr Thr Gln Glu Ala
 1285 1290 1295
 Asp Ser Ala Asn Thr Leu Gln Ile Ala Glu Ile Lys Gly Lys Ile Gly
 1300 1305 1310
 Thr Arg Ser Ala Glu Asp Pro Val Ser Glu Val Pro Ala Val Ser Gln
 1315 1320 1325
 His Pro Arg Thr Lys Ser Ser Arg Leu Gln Gly Ser Ser Leu Ser Ser
 1330 1335 1340
 Glu Ser Ala Arg His Lys Ala Val Glu Phe Pro Ser Gly Ala Lys Ser
 1345 1350 1355 1360
 Pro Ser Lys Ser Gly Ala Gln Thr Pro Lys Ser Pro Pro Glu His Tyr
 1365 1370 1375
 Val Gln Glu Thr Pro Leu Met Phe Ser Arg Cys Thr Ser Val Ser Ser
 1380 1385 1390
 Leu Asp Ser Phe Glu Ser Arg Ser Ile Ala Ser Ser Val Gln Ser Glu
 1395 1400 1405

Pro Cys Ser Gly Met Val Ser Gly Ile Ile Ser Pro Ser Asp Leu Pro
 1410 1415 1420
 Asp Ser Pro Gly Gln Thr Met Pro Pro Ser Arg Ser Lys Thr Pro Pro
 1425 1430 1435 1440
 Pro Pro Pro Gln Thr Ala Gln Thr Lys Arg Glu Val Pro Lys Asn Lys
 1445 1450 1455
 Ala Pro Thr Ala Glu Lys Arg Glu Ser Gly Pro Lys Gln Ala Ala Val
 1460 1465 1470
 Asn Ala Ala Val Gln Arg Val Gln Val Leu Pro Asp Ala Asp Thr Leu
 1475 1480 1485
 Leu His Phe Ala Thr Glu Ser Thr Pro Asp Gly Phe Ser Cys Ser Ser
 1490 1495 1500
 Ser Leu Ser Ala Leu Ser Leu Asp Glu Pro Phe Ile Gln Lys Asp Val
 1505 1510 1515 1520
 Glu Leu Arg Ile Met Pro Pro Val Gln Glu Asn Asp Asn Gly Asn Glu
 1525 1530 1535
 Thr Glu Ser Glu Gln Pro Lys Glu Ser Asn Glu Asn Gln Glu Lys Glu
 1540 1545 1550
 Ala Glu Lys Thr Ile Asp Ser Glu Lys Asp Leu Leu Asp Asp Ser Asp
 1555 1560 1565
 Asp Asp Asp Ile Glu Ile Leu Glu Glu Cys Ile Ile Ser Ala Met Pro
 1570 1575 1580
 Thr Lys Ser Ser Arg Lys Gly Lys Lys Pro Ala Gln Thr Ala Ser Lys
 1585 1590 1595 1600
 Leu Pro Pro Pro Val Ala Arg Lys Pro Ser Gln Leu Pro Val Tyr Lys
 1605 1610 1615
 Leu Leu Pro Ser Gln Asn Arg Leu Gln Pro Gln Lys His Val Ser Phe
 1620 1625 1630
 Thr Pro Gly Asp Asp Met Pro Arg Val Tyr Cys Val Glu Gly Thr Pro
 1635 1640 1645
 Ile Asn Phe Ser Thr Ala Thr Ser Leu Ser Asp Leu Thr Ile Glu Ser
 1650 1655 1660
 Pro Pro Asn Glu Leu Ala Ala Gly Glu Gly Val Arg Gly Gly Ala Gln
 1665 1670 1675 1680
 Ser Gly Glu Phe Glu Lys Arg Asp Thr Ile Pro Thr Glu Gly Arg Ser
 1685 1690 1695
 Thr Asp Glu Ala Gln Gly Gly Lys Thr Ser Ser Val Thr Ile Pro Glu
 1700 1705 1710
 Leu Asp Asp Asn Lys Ala Glu Glu Gly Asp Ile Leu Ala Glu Cys Ile
 1715 1720 1725
 Asn Ser Ala Met Pro Lys Gly Lys Ser His Lys Pro Phe Arg Val Lys
 1730 1735 1740

Lys Ile Met Asp Gln Val Gln Gln Ala Ser Ala Ser Ser Ser Ala Pro
 1745 1750 1755 1760
 Asn Lys Asn Gln Leu Asp Gly Lys Lys Lys Pro Thr Ser Pro Val
 1765 1770 1775
 Lys Pro Ile Pro Gln Asn Thr Glu Tyr Arg Thr Arg Val Arg Lys Asn
 1780 1785 1790
 Ala Asp Ser Lys Asn Asn Leu Asn Ala Glu Arg Val Phe Ser Asp Asn
 1795 1800 1805
 Lys Asp Ser Lys Lys Gln Asn Leu Lys Asn Asn Ser Lys Asp Phe Asn
 1810 1815 1820
 Asp Lys Leu Pro Asn Asn Glu Asp Arg Val Arg Gly Ser Phe Ala Phe
 1825 1830 1835 1840
 Asp Ser Pro His His Tyr Thr Pro Ile Glu Gly Thr Pro Tyr Cys Phe
 1845 1850 1855
 Ser Arg Asn Asp Ser Leu Ser Ser Leu Asp Phe Asp Asp Asp Val
 1860 1865 1870
 Asp Leu Ser Arg Glu Lys Ala Glu Leu Arg Lys Ala Lys Glu Asn Lys
 1875 1880 1885
 Glu Ser Glu Ala Lys Val Thr Ser His Thr Glu Leu Thr Ser Asn Gln
 1890 1895 1900
 Gln Ser Ala Asn Lys Thr Gln Ala Ile Ala Lys Gln Pro Ile Asn Arg
 1905 1910 1915 1920
 Gly Gln Pro Lys Pro Ile Leu Gln Lys Gln Ser Thr Phe Pro Gln Ser
 1925 1930 1935
 Ser Lys Asp Ile Pro Asp Arg Gly Ala Ala Thr Asp Glu Lys Leu Gln
 1940 1945 1950
 Asn Phe Ala Ile Glu Asn Thr Pro Val Cys Phe Ser His Asn Ser Ser
 1955 1960 1965
 Leu Ser Ser Leu Ser Asp Ile Asp Gln Glu Asn Asn Asn Lys Glu Asn
 1970 1975 1980
 Glu Pro Ile Lys Glu Thr Glu Pro Pro Asp Ser Gln Gly Glu Pro Ser
 1985 1990 1995 2000
 Lys Pro Gln Ala Ser Gly Tyr Ala Pro Lys Ser Phe His Val Glu Asp
 2005 2010 2015
 Thr Pro Val Cys Phe Ser Arg Asn Ser Ser Leu Ser Ser Leu Ser Ile
 2020 2025 2030
 Asp Ser Glu Asp Asp Leu Leu Gln Glu Cys Ile Ser Ser Ala Met Pro
 2035 2040 2045
 Lys Lys Lys Lys Pro Ser Arg Leu Lys Gly Asp Asn Glu Lys His Ser
 2050 2055 2060
 Pro Arg Asn Met Gly Gly Ile Leu Gly Glu Asp Leu Thr Leu Asp Leu
 2065 2070 2075 2080

Lys Asp Ile Gln Arg Pro Asp Ser Glu His Gly Leu Ser Pro Asp Ser
 2085 2090 2095
 Glu Asn Phe Asp Trp Lys Ala Ile Gln Glu Gly Ala Asn Ser Ile Val
 2100 2105 2110
 Ser Ser Leu His Gln Ala Ala Ala Ala Cys Leu Ser Arg Gln Ala
 2115 2120 2125
 Ser Ser Asp Ser Asp Ser Ile Leu Ser Leu Lys Ser Gly Ile Ser Leu
 2130 2135 2140
 Gly Ser Pro Phe His Leu Thr Pro Asp Gln Glu Glu Lys Pro Phe Thr
 2145 2150 2155 2160
 Ser Asn Lys Gly Pro Arg Ile Leu Lys Pro Gly Glu Lys Ser Thr Leu
 2165 2170 2175
 Glu Thr Lys Lys Ile Glu Ser Glu Ser Lys Gly Ile Lys Gly Gly Lys
 2180 2185 2190
 Lys Val Tyr Lys Ser Leu Ile Thr Gly Lys Val Arg Ser Asn Ser Glu
 2195 2200 2205
 Ile Ser Gly Gln Met Lys Gln Pro Leu Gln Ala Asn Met Pro Ser Ile
 2210 2215 2220
 Ser Arg Gly Arg Thr Met Ile His Ile Pro Gly Val Arg Asn Ser Ser
 2225 2230 2235 2240
 Ser Ser Thr Ser Pro Val Ser Lys Lys Gly Pro Pro Leu Lys Thr Pro
 2245 2250 2255
 Ala Ser Lys Ser Pro Ser Glu Gly Gln Thr Ala Thr Thr Ser Pro Arg
 2260 2265 2270
 Gly Ala Lys Pro Ser Val Lys Ser Glu Leu Ser Pro Val Ala Arg Gln
 2275 2280 2285
 Thr Ser Gln Ile Gly Gly Ser Ser Lys Ala Pro Ser Arg Ser Gly Ser
 2290 2295 2300
 Arg Asp Ser Thr Pro Ser Arg Pro Ala Gln Gln Pro Leu Ser Arg Pro
 2305 2310 2315 2320
 Ile Gln Ser Pro Gly Arg Asn Ser Ile Ser Pro Gly Arg Asn Gly Ile
 2325 2330 2335
 Ser Pro Pro Asn Lys Leu Ser Gln Leu Pro Arg Thr Ser Ser Pro Ser
 2340 2345 2350
 Thr Ala Ser Thr Lys Ser Ser Gly Ser Gly Lys Met Ser Tyr Thr Ser
 2355 2360 2365
 Pro Gly Arg Gln Met Ser Gln Gln Asn Leu Thr Lys Gln Thr Gly Leu
 2370 2375 2380
 Ser Lys Asn Ala Ser Ser Ile Pro Arg Ser Glu Ser Ala Ser Lys Gly
 2385 2390 2395 2400
 Leu Asn Gln Met Asn Asn Gly Asn Gly Ala Asn Lys Lys Val Glu Leu
 2405 2410 2415

Ser Arg Met Ser Ser Thr Lys Ser Ser Gly Ser Glu Ser Asp Arg Ser
 2420 2425 2430
 Glu Arg Pro Val Leu Val Arg Gln Ser Thr Phe Ile Lys Glu Ala Pro
 2435 2440 2445
 Ser Pro Thr Leu Arg Arg Lys Leu Glu Glu Ser Ala Ser Phe Glu Ser
 2450 2455 2460
 Leu Ser Pro Ser Ser Arg Pro Ala Ser Pro Thr Arg Ser Gln Ala Gln
 2465 2470 2475 2480
 Thr Pro Val Leu Ser Pro Ser Leu Pro Asp Met Ser Leu Ser Thr His
 2485 2490 2495
 Ser Ser Val Gln Ala Gly Gly Trp Arg Lys Leu Pro Pro Asn Leu Ser
 2500 2505 2510
 Pro Thr Ile Glu Tyr Asn Asp Gly Arg Pro Ala Lys Arg His Asp Ile
 2515 2520 2525
 Ala Arg Ser His Ser Glu Ser Pro Ser Arg Leu Pro Ile Asn Arg Ser
 2530 2535 2540
 Gly Thr Trp Lys Arg Glu His Ser Lys His Ser Ser Ser Leu Pro Arg
 2545 2550 2555 2560
 Val Ser Thr Trp Arg Arg Thr Gly Ser Ser Ser Ser Ile Leu Ser Ala
 2565 2570 2575
 Ser Ser Glu Ser Ser Glu Lys Ala Lys Ser Glu Asp Glu Lys His Val
 2580 2585 2590
 Asn Ser Ile Ser Gly Thr Lys Gln Ser Lys Glu Asn Gln Val Ser Ala
 2595 2600 2605
 Lys Gly Thr Trp Arg Lys Ile Lys Glu Asn Glu Phe Ser Pro Thr Asn
 2610 2615 2620
 Ser Thr Ser Gln Thr Val Ser Ser Gly Ala Thr Asn Gly Ala Glu Ser
 2625 2630 2635 2640
 Lys Thr Leu Ile Tyr Gln Met Ala Pro Ala Val Ser Lys Thr Glu Asp
 2645 2650 2655
 Val Trp Val Arg Ile Glu Asp Cys Pro Ile Asn Asn Pro Arg Ser Gly
 2660 2665 2670
 Arg Ser Pro Thr Gly Asn Thr Pro Pro Val Ile Asp Ser Val Ser Glu
 2675 2680 2685
 Lys Ala Asn Pro Asn Ile Lys Asp Ser Lys Asp Asn Gln Ala Lys Gln
 2690 2695 2700
 Asn Val Gly Asn Gly Ser Val Pro Met Arg Thr Val Gly Leu Glu Asn
 2705 2710 2715 2720
 Arg Leu Thr Ser Phe Ile Gln Val Asp Ala Pro Asp Gln Lys Gly Thr
 2725 2730 2735
 Glu Ile Lys Pro Gly Gln Asn Asn Pro Val Pro Val Ser Glu Thr Asn
 2740 2745 2750

Glu Ser Pro Ile Val Glu Arg Thr Pro Phe Ser Ser Ser Ser Ser Ser
 2755 2760 2765
 Lys His Ser Ser Pro Ser Gly Thr Val Ala Ala Arg Val Thr Pro Phe
 2770 2775 2780
 Asn Tyr Asn Pro Ser Pro Arg Lys Ser Ser Ala Asp Ser Thr Ser Ala
 2785 2790 2795 2800
 Arg Pro Ser Gln Ile Pro Thr Pro Val Asn Asn Asn Thr Lys Lys Arg
 2805 2810 2815
 Asp Ser Lys Thr Asp Ser Thr Glu Ser Ser Gly Thr Gln Ser Pro Lys
 2820 2825 2830
 Arg His Ser Gly Ser Tyr Leu Val Thr Ser Val
 2835 2840

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3172 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
 (B) CLONE: DP1(T82)

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 1..630

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCA GTC GCC GCT CCA GTC TAT GCG GCA CTA GGA ACA GCG CCG CGA GCG	48
Ala Val Ala Ala Pro Val Tyr Pro Ala Leu Gly Thr Ala Pro Gly Gly	15
1 5 10	
GAG ACC GTC CCC GCC ATG TCT GCG GCC ATG ACC GAG ACC TTC GAC CCG	96
Glu Thr Val Pro Ala Met Ser Ala Ala Met Arg Glu Arg Phe Asp Arg	30
20 25 30	
TTC CTG CAC CAG AAG AAC TGC ATG ACT GAC CTT CTG GCC AAG CTC CAG	144
Phe Leu His Glu Lys Asn Cys Met Thr Asp Leu Leu Ala Lys Leu Glu	45
35 40 45	
GCC AAA ACC GCG GTG AAC ACC ACC TTC ATC GCT GTT GGT GTC ATC GGA	192
Ala Lys Thr Gly Val Asn Arg Ser Phe Ile Ala Leu Gly Val Ile Gly	60
50 55 60	
CTG GTG GCC TTG TAC CTG CTC TTC GGT TAT GGA GCC TCT CTC CTC TGC	240
Leu Val Ala Leu Tyr Leu Val Phe Gly Tyr Gly Ala Ser Leu Leu Cys	80
65 70 75 80	

AAC CTC ATA CGA TTT GGC TAC CCA GCC TAC ATC TCA ATT AAA GCT ATA Asn Leu Ile Gly Phe Gly Tyr Pro Ala Tyr Ile Ser Ile Lys Ala Ile 85 90 95	288
CAG AGT CCC AAC AAA GAA GAT CAT ACC CAG TGG CTG ACC TAC TGG GTA Glu Ser Pro Asn Lys Glu Asp Asp Thr Gln Trp Leu Thr Tyr Trp Val 100 105 110	336
CTG TAT GGT GTG TTC AGC ATT GCT GAA TTC TTC TCT GAT ATC TTC CTG Val Tyr Gly Val Phe Ser Ile Ala Glu Phe Phe Ser Asp Ile Phe Leu 115 120 125	384
TCA TGG TTC CCC TTC TAC TAC ATC CTC AAC TGT GCC TTC CTG TTG TGG Ser Trp Phe Pro Phe Tyr Tyr Met Leu Lys Cys Gly Phe Leu Leu Trp 130 135 140	432
TGC ATG GCC CCG AGC CCT TCT AAT GCG CCT GAA CTG CTC TAC AAG CGC Cys Met Ala Pro Ser Pro Ser Asn Gly Ala Glu Leu Leu Tyr Lys Arg 145 150 155 160	480
ATC ATC CGT CCT TTC TTC CTG AAG CAC GAG TCC CAC ATC GAC AGT GTG Ile Ile Arg Pro Phe Phe Leu Lys His Glu Ser Gln Met Asp Ser Val 165 170 175	528
CTC AAG GAC CTT AAA GAC AAG TCC AAA GAG ACT GCA GAT GCC ATC ACT Val Lys Asp Leu Lys Asp Lys Ser Lys Glu Thr Ala Asp Ala Ile Thr 180 185 190	576
AAA GAA GCC AAG AAA GCT ACC GTG AAT TTA CTC GGT GAA GAA AAG AAG Lys Glu Ala Lys Lys Ala Thr Val Asn Leu Leu Gly Glu Glu Lys Lys 195 200 205	624
AGC ACC TAAACCAGAC TAAACCAGAC TGCATCGAAA GTTCCTGCCC TCTCTGTACC Ser Thr 210	680
ITCTTACTGG ACCTTGATGT TATATTAGGG ACTGTGGTAT AATTATTTTA ATAATGTTGC	740
CTTGGAAACA TTTTTCAGAT ATTAAGATT GGATGTGTT GTAAGTTTCT TTGCTTACTT	800
TTACTGTCTA TATATATAGC CACCACCTTA AACTTAATGC AGTGGGCAGT GTCCACGTTT	860
TTGGAAATG TATTTTGCCT CTGGGTAGGA AAAGATGTAT GTTGTATCC TGCAGGAAAT	920
ATAAACTTAA AATAAAATTA TATACCCCAAC AGGCTGTGTA CTTTACTGGG CTCTCCCTGC	980
ACGSATTTTC TCTGTAGTTA CATTAGGRT AATCTTATG GTTCTACTTC CTATAATGTA	1040
CAATTTTATA TAATTCTGTA ATGTTTTTAA TGTATTTGTC CACATGTACA TATGGAAATG	1100
TTACTGTCTG ACTACANCAAT GCATCATGCT CATGGGGAGG CAGCAGCGGA AGCTTGTATG	1160
TGTCATTIAT AACTTCTGTA CAGTAAGACC ACCTGCCAAA AGCTGGAGGA ACCATTGTGC	1220
TGCTGTGGTC TACTAAATAA TACTTTAGGA AATACGTGAT TAATATGCAA GTGAACAAAG	1280
TCAGAAATGA AATCGAATGG ACATTGGCCT GCTTGTITCC GTAGTATATG GCATATGAAT	1340
ACCAGCATAG CTTTATAAAG CAGTTAGTTA GTTAGTTACT CACTCTAGTG ATAAATCGGG	1400
AAATTTACAC ACACACACAC ACACACACAC ACACACACAC ACACACACAC ACACACACAG	1460

AGTACCCCTGT AACTCTCAAT TCCCTGAAAA ACTAGTAATA GTGTCTTATC TGCTATAAAC	1520
TTTACATAAT TGTCTATTGT CAAGATGCTA CANTGGAMNC CATTTCTGCT TTTATCTTCA	1580
NAGSGGAGAN ACATGTTGAT TTAGTCTTCT TTCCCAATCT TCTTTTTTAA MCCAGTTTNA	1640
GGHNGTTCTG RAGATTGTC CACCTCTGAT TACATCTATC TTCTTGTTTG TATCATKACC	1700
AACACATGC TAATGRCGAC ACCTAGCTCT RAGHGCCAAT CTGGGACANT GARAGGNWGT	1760
ATAPAGTHNC CCATAATCTG GTTGGCAATA GTTAAGTCAA TCTATCTTCA GTTTTTCTCT	1820
CCCCTTTAAG GTCAACACA ACAGGCTTCC CTAGTTTACA AGTCAGAGTC ACTTGAGTC	1880
CATTTAAATG CCGTCATCCG TATTCTTTGT GTTGATNAGC TGCACAKGAC TACATAGTAA	1940
GTACAGANCA GTAAAGTTAA NNCOGATGTC TCCATTGATC TGCCAAHTCC NTATAGAGAG	2000
CAATTGTCT GCACTAGAAA ATCTGAGTTT TACACCATAC TGTTAAGAGT CCTTTTGAAT	2060
TAACTAGAC TAAACAACT GTATAACTAA ACTAACAGA TTAATATCC AGCCAGTACA	2120
GTATTTTTTA AGCCAAATA AGATGATTAG CTCACCTTGA GHTAACAAATC AGCTAAGATC	2180
ATNACAATGT CTCATGATGT NAANAATATT AAGCATATCA ATACTAAGTC ACAGTATCAC	2240
MNCTAATATA ATATGATCA GAGCATTTAT TTGCGGAGC AAAACAGTGG TGATTACCGG	2300
CATTTTATTA AACTAAAC TTTGTAGAAA GCAACAAAA TTGTTCTTGG GAGAAAAACA	2360
ACTTTTAGAT TAAAAAAT TTAAGTANCT AGGAGTATTT AATCCTTTT CCCATAAATA	2420
AAAGTACAST TTTCTTGGTG GCAGAAATGA AATCAGCAAC NTCTAGCATA TACAGTATAT	2480
AATCAGATTG ACAGCATATA GAATATATTA TCAGACAAAG TGAGGAGGTA CAAAAGTTAC	2540
TATTGCTCAT AATGACTTAC AGGCTAAAA TAGNTATAA ATACTATAIT AATTTCTGAA	2600
TGCAATTTTT TTTGTTCCC TTGAGACCAA AATTAAAGTT AACTCTTGGT GGCAGTCTAA	2660
GTGTAAATGT TAAACGACG AGAAGTTAAG AATGAGCAG TTCTGTGCA TGATTTCCTA	2720
AATGAATAAC TGCCTTGGCT AGAGTTTGAA AACTAATTG AGCCTGTGCC TGCCTAGAAA	2780
ACAGCGTTT ATTGAATGT GAATAGTGT TCAAGGTAT GTAGTTACAG AATTCCTACC	2840
AAACAGCTTA AATTGTTCAA GAAAGAAATC CTCAGCAGT TATTCCTTA CCTGAAGGCT	2900
TCAATCATTT GGATCAACAA CTGCTACTCT CGGGAAGAC CCTCTACTCA CAGCTGAAGA	2960
AAATGAGCAG ACCCTTCACA CTGTTATCAC CTATCCTGAA CATGTGATAC ACTGAATGGA	3020
AATAAATAGA TGTAAATAAA ATTGAGWTCT CATTTAAAAA AAACCATGTC CCCAATGCGA	3080
AAATGACCTC ATGTTCTGCT TTAACAGCA ACTGCACCCA CTAGCACAGC CCATTGAGCT	3140
ANCCTATATA TACATCTCTG TCAGTCCCCC TC	3172

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 210 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Ala Val Ala Ala Pro Val Tyr Pro Ala Leu Gly Thr Ala Pro Gly Gly
 1           5           10           15
Glu Thr Val Pro Ala Met Ser Ala Ala Met Arg Glu Arg Phe Asp Arg
 20           25           30
Phe Leu His Glu Lys Asn Cys Met Thr Asp Leu Leu Ala Lys Leu Glu
 35           40           45
Ala Lys Thr Gly Val Asn Arg Ser Phe Ile Ala Leu Gly Val Ile Gly
 50           55           60
Leu Val Ala Leu Tyr Leu Val Phe Gly Tyr Gly Ala Ser Leu Leu Cys
 65           70           75           80
Asn Leu Ile Gly Phe Gly Tyr Pro Ala Tyr Ile Ser Ile Lys Ala Ile
 85           90           95
Glu Ser Pro Asn Lys Glu Asp Asp Thr Gln Trp Leu Thr Tyr Trp Val
100           105           110
Val Tyr Gly Val Phe Ser Ile Ala Glu Phe Phe Ser Asp Ile Phe Leu
115           120           125
Ser Trp Phe Pro Phe Tyr Tyr Met Leu Lys Cys Gly Phe Leu Leu Trp
130           135           140
Cys Met Ala Pro Ser Pro Ser Asn Gly Ala Glu Leu Leu Tyr Lys Arg
145           150           155           160
Ile Ile Arg Pro Phe Phe Leu Lys His Glu Ser Gln Met Asp Ser Val
165           170           175
Val Lys Asp Leu Lys Asp Lys Ser Lys Glu Thr Ala Asp Ala Ile Thr
180           185           190
Lys Glu Ala Lys Lys Ala Thr Val Asn Leu Leu Gly Glu Glu Lys Lys
195           200           205
Ser Thr
210

```

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 434 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
(B) CLONE: T81

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```

Val Ala Pro Val Val Val Gly Ser Gly Arg Ala Pro Arg His Pro Ala
 1              5              10              15
Pro Ala Ala Met His Pro Arg Arg Pro Asp Gly Phe Asp Gly Leu Gly
 20              25              30
Tyr Arg Gly Gly Ala Arg Asp Glu Gln Gly Phe Gly Gly Ala Phe Pro
 35              40              45
Ala Arg Ser Phe Ser Thr Gly Ser Asp Leu Gly His Trp Val Thr Thr
 50              55              60
Pro Pro Asp Ile Pro Gly Ser Arg Asn Leu His Trp Gly Glu Lys Ser
 65              70              75              80
Pro Pro Tyr Gly Val Pro Thr Thr Ser Thr Pro Tyr Glu Gly Pro Thr
 85              90              95
Glu Glu Pro Phe Ser Ser Gly Gly Gly Gly Ser Val Gln Gly Gln Ser
100              105              110
Ser Glu Gln Leu Asn Arg Phe Ala Gly Phe Gly Ile Gly Leu Ala Ser
115              120              125
Leu Phe Thr Glu Asn Val Leu Ala His Pro Cys Ile Val Leu Arg Arg
130              135              140
Gln Cys Gln Val Asn Tyr His Ala Gln His Tyr His Leu Thr Pro Phe
145              150              155              160
Thr Val Ile Asn Ile Met Tyr Ser Phe Asn Lys Thr Gln Gly Pro Arg
165              170              175
Ala Leu Trp Lys Gly Met Gly Ser Thr Phe Ile Val Gln Gly Val Thr
180              185              190
Leu Gly Ala Glu Gly Ile Ile Ser Glu Phe Thr Pro Leu Pro Arg Glu
195              200              205
Val Leu His Lys Trp Ser Pro Lys Gln Ile Gly Glu His Leu Leu Leu
210              215              220
Lys Ser Leu Thr Tyr Val Val Ala Met Pro Phe Tyr Ser Ala Ser Leu
225              230              235              240
Ile Glu Thr Val Gln Ser Glu Ile Ile Arg Asp Asn Thr Gly Ile Leu
245              250              255
Glu Cys Val Lys Glu Gly Ile Gly Arg Val Ile Gly Met Gly Val Pro
260              265              270

```

His Ser Lys Arg Leu Leu Pro Leu Leu Ser Leu Ile Phe Pro Thr Val
 275 280 285
 Leu His Gly Val Leu His Tyr Ile Ile Ser Ser Val Ile Gln Lys Phe
 290 295 300
 Val Leu Leu Ile Leu Lys Arg Lys Thr Tyr Asn Ser His Leu Ala Glu
 305 310 315 320
 Ser Thr Ser Pro Val Gln Ser Met Leu Asp Ala Tyr Phe Pro Glu Leu
 325 330 335
 Ile Ala Asn Phe Ala Ala Ser Leu Cys Ser Asp Val Ile Leu Tyr Pro
 340 345 350
 Leu Glu Thr Val Leu His Arg Leu His Ile Gln Gly Thr Arg Thr Ile
 355 360 365
 Ile Asp Asn Thr Asp Leu Gly Tyr Glu Val Leu Pro Ile Asn Thr Gln
 370 375 380
 Tyr Glu Gly Met Arg Asp Cys Ile Asn Thr Ile Arg Gln Glu Glu Gly
 385 390 395 400
 Val Phe Gly Phe Tyr Lys Gly Phe Gly Ala Val Ile Ile Gln Tyr Thr
 405 410 415
 Leu His Ala Ala Val Leu Gln Ile Thr Lys Ile Ile Tyr Ser Thr Leu
 420 425 430
 Leu Gln

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 185 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:

- (B) CLONE: YS-39(T82)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Glu Leu Arg Arg Phe Asp Arg Phe Leu His Glu Lys Asn Cys Met Thr
 1 5 10 15
 Asp Leu Leu Ala Lys Leu Glu Ala Lys Thr Gly Val Asn Arg Ser Phe
 20 25 30
 Ile Ala Leu Gly Val Ile Gly Leu Val Ala Leu Tyr Leu Val Phe Gly
 35 40 45

Tyr Gly Ala Ser Leu Leu Cys Asn Leu Ile Gly Phe Gly Tyr Pro Ala
 50 55 60
 Tyr Ile Ser Ile Lys Ala Ile Glu Ser Pro Asn Lys Glu Asp Asp Thr
 65 70 75 80
 Gln Trp Leu Thr Tyr Trp Val Val Tyr Gly Val Phe Ser Ile Ala Glu
 85 90 95
 Phe Phe Ser Asp Ile Phe Leu Ser Trp Phe Pro Phe Tyr Tyr Ile Leu
 100 105 110
 Lys Cys Gly Phe Leu Leu Trp Cys Met Ala Pro Ser Pro Ser Asn Gly
 115 120 125
 Ala Glu Leu Leu Tyr Lys Arg Ile Ile Arg Pro Phe Phe Leu Lys His
 130 135 140
 Glu Ser Gln Met Asp Ser Val Val Lys Asp Leu Lys Asp Lys Ala Lys
 145 150 155 160
 Glu Thr Ala Asp Ala Ile Thr Lys Glu Ala Lys Lys Ala Thr Val Asn
 165 170 175
 Leu Leu Gly Glu Glu Lys Lys Ser Thr
 180 185

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2842 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:

- (B) CLONE: APC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Ala Ala Ala Ser Tyr Asp Gln Leu Leu Lys Gln Val Glu Ala Leu
 1 5 10 15
 Lys Met Glu Asn Ser Asn Leu Arg Gln Glu Leu Glu Asp Asn Ser Asn
 20 25 30
 His Leu Thr Lys Leu Glu Thr Glu Ala Ser Asn Met Lys Glu Val Leu
 35 40 45
 Lys Gln Leu Gln Gly Ser Ile Glu Asp Glu Ala Met Ala Ser Ser Gly
 50 55 60
 Gln Ile Asp Leu Leu Glu Arg Leu Lys Glu Leu Asn Leu Asp Ser Ser
 65 70 75 80

Asn Phe Pro Gly Val Lys Leu Arg Ser Lys Met Ser Leu Arg Ser Tyr
 85 90 95
 Gly Ser Arg Glu Gly Ser Val Ser Ser Arg Ser Gly Glu Cys Ser Pro
 100 105 110
 Val Pro Met Gly Ser Phe Pro Arg Arg Gly Phe Val Asn Gly Ser Arg
 115 120 125
 Glu Ser Thr Gly Tyr Leu Glu Glu Leu Glu Lys Glu Arg Ser Leu Leu
 130 135 140
 Leu Ala Asp Leu Asp Lys Glu Glu Lys Glu Lys Asp Trp Tyr Tyr Ala
 145 150 155 160
 Gln Leu Gln Asn Leu Thr Lys Arg Ile Asp Ser Leu Leu Thr Glu Asn
 165 170 175
 Phe Ser Leu Gln Thr Asp Met Thr Arg Arg Gln Leu Glu Tyr Glu Ala
 180 185 190
 Arg Gln Ile Arg Val Ala Met Glu Glu Gln Leu Gly Thr Cys Gln Asp
 195 200 205
 Met Glu Lys Arg Ala Gln Arg Arg Ile Ala Arg Ile Gln Gln Ile Glu
 210 215 220
 Lys Asp Ile Leu Arg Ile Arg Gln Leu Leu Gln Ser Gln Ala Thr Glu
 225 230 235 240
 Ala Glu Arg Ser Ser Gln Asn Lys His Glu Thr Gly Ser His Asp Ala
 245 250 255
 Glu Arg Gln Asn Glu Gly Gln Gly Val Gly Glu Ile Asn Met Ala Thr
 260 265 270
 Ser Gly Asn Gly Gln Gly Ser Thr Thr Arg Met Asp His Glu Thr Ala
 275 280 285
 Ser Val Leu Ser Ser Ser Ser Thr His Ser Ala Pro Arg Arg Leu Thr
 290 295 300
 Ser His Leu Gly Thr Lys Val Glu Met Val Tyr Ser Leu Leu Ser Met
 305 310 315 320
 Leu Gly Thr His Asp Lys Asp Asp Met Ser Arg Thr Leu Leu Ala Met
 325 330 335
 Ser Ser Ser Gln Asp Ser Cys Ile Ser Met Arg Gln Ser Gly Cys Leu
 340 345 350
 Pro Leu Leu Ile Gln Leu Leu His Gly Asn Asp Lys Asp Ser Val Leu
 355 360 365
 Leu Gly Asn Ser Arg Gly Ser Lys Glu Ala Arg Ala Arg Ala Ser Ala
 370 375 380
 Ala Leu His Asn Ile Ile His Ser Gln Pro Asp Asp Lys Arg Gly Arg
 385 390 395 400
 Arg Glu Ile Arg Val Leu His Leu Leu Glu Gln Ile Arg Ala Tyr Cys
 405 410 415

Glu Thr Cys Trp Glu Trp Gln Glu Ala His Glu Pro Gly Met Asp Gln
 420 425 430
 Asp Lys Asn Pro Met Pro Ala Pro Val Glu His Gln Ile Cys Pro Ala
 435 440 445
 Val Cys Val Leu Met Lys Leu Ser Phe Asp Glu Glu His Arg His Ala
 450 455 460
 Met Asn Glu Leu Gly Gly Leu Gln Ala Ile Ala Glu Leu Leu Gln Val
 465 470 475 480
 Asp Cys Glu Met Tyr Gly Leu Thr Asn Asp His Tyr Ser Ile Thr Leu
 485 490 495
 Arg Arg Tyr Ala Gly Met Ala Leu Thr Asn Leu Thr Phe Gly Asp Val
 500 505 510
 Ala Asn Lys Ala Thr Leu Cys Ser Met Lys Gly Cys Met Arg Ala Leu
 515 520 525
 Val Ala Gln Leu Lys Ser Glu Ser Glu Asp Leu Gln Gln Val Ile Ala
 530 535 540
 Ser Val Leu Arg Asn Leu Ser Trp Arg Ala Asp Val Asn Ser Lys Lys
 545 550 555 560
 Thr Leu Arg Glu Val Gly Ser Val Lys Ala Leu Met Glu Cys Ala Leu
 565 570 575
 Glu Val Lys Lys Glu Ser Thr Leu Lys Ser Val Leu Ser Ala Leu Trp
 580 585 590
 Asn Leu Ser Ala His Cys Thr Glu Asn Lys Ala Asp Ile Cys Ala Val
 595 600 605
 Asp Gly Ala Leu Ala Phe Leu Val Gly Thr Leu Thr Tyr Arg Ser Gln
 610 615 620
 Thr Asn Thr Leu Ala Ile Ile Glu Ser Gly Gly Gly Ile Leu Arg Asn
 625 630 635 640
 Val Ser Ser Leu Ile Ala Thr Asn Glu Asp His Arg Gln Ile Leu Arg
 645 650 655
 Glu Asn Asn Cys Leu Gln Thr Leu Leu Gln His Leu Lys Ser His Ser
 660 665 670
 Leu Thr Ile Val Ser Asn Ala Cys Gly Thr Leu Trp Asn Leu Ser Ala
 675 680 685
 Arg Asn Pro Lys Asp Gln Glu Ala Leu Trp Asp Met Gly Ala Val Ser
 690 695 700
 Met Leu Lys Asn Leu Ile His Ser Lys His Lys Met Ile Ala Met Gly
 705 710 715 720
 Ser Ala Ala Ala Leu Arg Asn Leu Met Ala Asn Arg Pro Ala Lys Tyr
 725 730 735
 Lys Asp Ala Asn Ile Met Ser Pro Gly Ser Ser Leu Pro Ser Leu His
 740 745 750

Val Arg Lys Gln Lys Ala Leu Glu Ala Glu Leu Asp Ala Gln His Leu
 755 760 765
 Ser Glu Thr Phe Asp Asn Ile Asp Asn Leu Ser Pro Lys Ala Ser His
 770 775 780
 Arg Ser Lys Gln Arg His Lys Gln Ser Leu Tyr Gly Asp Tyr Val Phe
 785 790 795 800
 Asp Thr Asn Arg His Asp Asp Asn Arg Ser Asp Asn Phe Asn Thr Gly
 805 810 815
 Asn Met Thr Val Leu Ser Pro Tyr Leu Asn Thr Thr Val Leu Pro Ser
 820 825 830
 Ser Ser Ser Ser Arg Gly Ser Leu Asp Ser Ser Arg Ser Glu Lys Asp
 835 840 845
 Arg Ser Leu Glu Arg Glu Arg Gly Ile Gly Leu Gly Asn Tyr His Pro
 850 855 860
 Ala Thr Glu Asn Pro Gly Thr Ser Ser Lys Arg Gly Leu Gln Ile Ser
 865 870 875 880
 Thr Thr Ala Ala Gln Ile Ala Lys Val Met Glu Glu Val Ser Ala Ile
 885 890 895
 His Thr Ser Gln Glu Asp Arg Ser Ser Gly Ser Thr Thr Glu Leu His
 900 905 910
 Cys Val Thr Asp Glu Arg Asn Ala Leu Arg Arg Ser Ser Ala Ala His
 915 920 925
 Thr His Ser Asn Thr Tyr Asn Phe Thr Lys Ser Glu Asn Ser Asn Arg
 930 935 940
 Thr Cys Ser Met Pro Tyr Ala Lys Leu Glu Tyr Lys Arg Ser Ser Asn
 945 950 955 960
 Asp Ser Leu Asn Ser Val Ser Ser Ser Asp Gly Tyr Gly Lys Arg Gly
 965 970 975
 Gln Met Lys Pro Ser Ile Glu Ser Tyr Ser Glu Asp Asp Glu Ser Lys
 980 985 990
 Phe Cys Ser Tyr Gly Gln Tyr Pro Ala Asp Leu Ala His Lys Ile His
 995 1000 1005
 Ser Ala Asn His Met Asp Asp Asn Asp Gly Glu Leu Asp Thr Pro Ile
 1010 1015 1020
 Asn Tyr Ser Leu Lys Tyr Ser Asp Glu Gln Leu Asn Ser Gly Arg Gln
 1025 1030 1035 1040
 Ser Pro Ser Gln Asn Glu Arg Trp Ala Arg Pro Lys His Ile Ile Glu
 1045 1050 1055
 Asp Glu Ile Lys Gln Ser Glu Gln Arg Gln Ser Arg Asn Gln Ser Thr
 1060 1065 1070
 Thr Tyr Pro Val Tyr Thr Glu Ser Thr Asp Asp Lys His Leu Lys Phe
 1075 1080 1085

Gln Pro His Phe Gly Gln Gln Glu Cys Val Ser Pro Tyr Arg Ser Arg
 1090 1095 1100
 Gly Ala Asn Gly Ser Glu Thr Asn Arg Val Gly Ser Asn His Gly Ile
 1105 1110 1115 1120
 Asn Gln Asn Val Ser Gln Ser Leu Cys Gln Glu Asp Asp Tyr Glu Asp
 1125 1130 1135
 Asp Lys Pro Thr Asn Tyr Ser Glu Arg Tyr Ser Glu Glu Glu Gln His
 1140 1145 1150
 Glu Glu Glu Glu Arg Pro Thr Asn Tyr Ser Ile Lys Tyr Asn Glu Glu
 1155 1160 1165
 Lys Arg His Val Asp Gln Pro Ile Asp Tyr Ser Leu Lys Tyr Ala Thr
 1170 1175 1180
 Asp Ile Pro Ser Ser Gln Lys Gln Ser Phe Ser Phe Ser Lys Ser Ser
 1185 1190 1195 1200
 Ser Gly Gln Ser Ser Lys Thr Glu His Met Ser Ser Ser Ser Glu Asn
 1205 1210 1215
 Thr Ser Thr Pro Ser Ser Asn Ala Lys Arg Gln Asn Gln Leu His Pro
 1220 1225 1230
 Ser Ser Ala Gln Ser Arg Ser Gly Gln Pro Gln Lys Ala Ala Thr Cys
 1235 1240 1245
 Lys Val Ser Ser Ile Asn Gln Glu Thr Ile Gln Thr Tyr Cys Val Glu
 1250 1255 1260
 Asp Thr Pro Ile Cys Phe Ser Arg Cys Ser Ser Leu Ser Ser Leu Ser
 1265 1270 1275 1280
 Ser Ala Glu Asp Glu Ile Gly Cys Asn Gln Thr Thr Gln Glu Ala Asp
 1285 1290 1295
 Ser Ala Asn Thr Leu Gln Ile Ala Glu Ile Lys Glu Lys Ile Gly Thr
 1300 1305 1310
 Arg Ser Ala Glu Asp Pro Val Ser Glu Val Pro Ala Val Ser Gln His
 1315 1320 1325
 Pro Arg Thr Lys Ser Ser Arg Leu Gln Gly Ser Ser Leu Ser Ser Glu
 1330 1335 1340
 Ser Ala Arg His Lys Ala Val Glu Phe Ser Ser Gly Ala Lys Ser Pro
 1345 1350 1355 1360
 Ser Lys Ser Gly Ala Gln Thr Pro Lys Ser Pro Pro Glu His Tyr Val
 1365 1370 1375
 Gln Glu Thr Pro Leu Met Phe Ser Arg Cys Thr Ser Val Ser Ser Leu
 1380 1385 1390
 Asp Ser Phe Glu Ser Arg Ser Ile Ala Ser Ser Val Gln Ser Glu Pro
 1395 1400 1405
 Cys Ser Gly Met Val Ser Gly Ile Ile Ser Pro Ser Asp Leu Pro Asp
 1410 1415 1420

Ser Pro Gly Gln Thr Met Pro Pro Ser Arg Ser Lys Thr Pro Pro Pro
 1425 1430 1435 1440
 Pro Pro Gln Thr Ala Gln Thr Lys Arg Glu Val Pro Lys Asn Lys Ala
 1445 1450 1455
 Pro Thr Ala Glu Lys Arg Glu Ser Gly Pro Lys Gln Ala Ala Val Asn
 1460 1465 1470
 Ala Ala Val Gln Arg Val Gln Val Leu Pro Asp Ala Asp Thr Leu Leu
 1475 1480 1485
 His Phe Ala Thr Glu Ser Thr Pro Asp Gly Phe Ser Cys Ser Ser Ser
 1490 1495 1500
 Leu Ser Ala Leu Ser Leu Asp Glu Pro Phe Ile Gln Lys Asp Val Glu
 1505 1510 1515 1520
 Leu Arg Ile Met Pro Pro Val Gln Glu Asn Asp Asn Gly Asn Glu Thr
 1525 1530 1535
 Glu Ser Glu Gln Pro Lys Glu Ser Asn Glu Asn Gln Glu Lys Glu Ala
 1540 1545 1550
 Glu Lys Thr Ile Asp Ser Glu Lys Asp Leu Leu Asp Asp Ser Asp Asp
 1555 1560 1565
 Asp Asp Ile Glu Ile Leu Glu Glu Cys Ile Ile Ser Ala Met Pro Thr
 1570 1575 1580
 Lys Ser Ser Arg Lys Ala Lys Lys Pro Ala Gln Thr Ala Ser Lys Leu
 1585 1590 1595 1600
 Pro Pro Pro Val Ala Arg Lys Pro Ser Gln Leu Pro Val Tyr Lys Leu
 1605 1610 1615
 Leu Pro Ser Gln Asn Arg Leu Gln Pro Gln Lys His Val Ser Phe Thr
 1620 1625 1630
 Pro Gly Asp Asp Met Pro Arg Val Tyr Cys Val Glu Gly Thr Pro Ile
 1635 1640 1645
 Asn Phe Ser Thr Ala Thr Ser Leu Ser Asp Leu Thr Ile Glu Ser Pro
 1650 1655 1660
 Pro Asn Glu Leu Ala Ala Gly Glu Gly Val Arg Gly Gly Ala Gln Ser
 1665 1670 1675 1680
 Gly Glu Phe Glu Lys Arg Asp Thr Ile Pro Thr Glu Gly Arg Ser Thr
 1685 1690 1695
 Asp Glu Ala Gln Gly Gly Lys Thr Ser Ser Val Thr Ile Pro Glu Leu
 1700 1705 1710
 Asp Asp Asn Lys Ala Glu Glu Gly Asp Ile Leu Ala Glu Cys Ile Asn
 1715 1720 1725
 Ser Ala Met Pro Lys Gly Lys Ser His Lys Pro Phe Arg Val Lys Lys
 1730 1735 1740
 Ile Met Asp Gln Val Gln Gln Ala Ser Ala Ser Ser Ser Ala Pro Asn
 1745 1750 1755 1760

Lys Asn Gln Leu Asp Gly Lys Lys Lys Lys Pro Thr Ser Pro Val Lys
 1765 1770 1775
 Pro Ile Pro Gln Asn Thr Glu Tyr Arg Thr Arg Val Arg Lys Asn Ala
 1780 1785 1790
 Asp Ser Lys Asn Asn Leu Asn Ala Glu Arg Val Phe Ser Asp Asn Lys
 1795 1800 1805
 Asp Ser Lys Lys Gln Asn Leu Lys Asn Asn Ser Lys Asp Phe Asn Asp
 1810 1815 1820
 Lys Leu Pro Asn Asn Glu Asp Arg Val Arg Gly Ser Phe Ala Phe Asp
 1825 1830 1835 1840
 Ser Pro His His Tyr Thr Pro Ile Glu Gly Thr Pro Tyr Cys Phe Ser
 1845 1850 1855
 Arg Asn Asp Ser Leu Ser Ser Leu Asp Phe Asp Asp Asp Val Asp
 1860 1865 1870
 Leu Ser Arg Glu Lys Ala Glu Leu Arg Lys Ala Lys Glu Asn Lys Glu
 1875 1880 1885
 Ser Glu Ala Lys Val Thr Ser His Thr Glu Leu Thr Ser Asn Gln Gln
 1890 1895 1900
 Ser Ala Asn Lys Thr Gln Ala Ile Ala Lys Gln Pro Ile Asn Arg Gly
 1905 1910 1915 1920
 Gln Pro Lys Pro Ile Leu Gln Lys Gln Ser Thr Phe Pro Gln Ser Ser
 1925 1930 1935
 Lys Asp Ile Pro Asp Arg Gly Ala Ala Thr Asp Glu Lys Leu Gln Asn
 1940 1945 1950
 Phe Ala Ile Glu Asn Thr Pro Val Cys Phe Ser His Asn Ser Ser Leu
 1955 1960 1965
 Ser Ser Leu Ser Asp Ile Asp Gln Glu Asn Asn Asn Lys Glu Asn Glu
 1970 1975 1980
 Pro Ile Lys Glu Thr Glu Pro Pro Asp Ser Gln Gly Glu Pro Ser Lys
 1985 1990 1995 2000
 Pro Gln Ala Ser Gly Tyr Ala Pro Lys Ser Phe His Val Glu Asp Thr
 2005 2010 2015
 Pro Val Cys Phe Ser Arg Asn Ser Ser Leu Ser Ser Leu Ser Ile Asp
 2020 2025 2030
 Ser Glu Asp Asp Leu Leu Gln Glu Cys Ile Ser Ser Ala Met Pro Lys
 2035 2040 2045
 Lys Lys Lys Pro Ser Arg Leu Lys Gly Asp Asn Glu Lys His Ser Pro
 2050 2055 2060
 Arg Asn Met Gly Gly Ile Leu Gly Glu Asp Leu Thr Leu Asp Leu Lys
 2065 2070 2075 2080
 Asp Ile Gln Arg Pro Asp Ser Glu His Gly Leu Ser Pro Asp Ser Glu
 2085 2090 2095

Asn Phe Asp Trp Lys Ala Ile Gln Glu Gly Ala Asn Ser Ile Val Ser
 2100 2103 2110
 Ser Leu His Gln Ala Ala Ala Ala Cys Leu Ser Arg Gln Ala Ser
 2115 2120 2125
 Ser Asp Ser Asp Ser Ile Leu Ser Leu Lys Ser Gly Ile Ser Leu Gly
 2130 2135 2140
 Ser Pro Phe His Leu Thr Pro Asp Gln Glu Glu Lys Pro Phe Thr Ser
 2145 2150 2155 2160
 Asn Lys Gly Pro Arg Ile Leu Lys Pro Gly Glu Lys Ser Thr Leu Glu
 2165 2170 2175
 Thr Lys Lys Ile Glu Ser Glu Ser Lys Gly Ile Lys Gly Lys Lys
 2180 2185 2190
 Val Tyr Lys Ser Leu Ile Thr Gly Lys Val Arg Ser Asn Ser Glu Ile
 2195 2200 2205
 Ser Gly Gln Met Lys Gln Pro Leu Gln Ala Asn Met Pro Ser Ile Ser
 2210 2215 2220
 Arg Gly Arg Thr Met Ile His Ile Pro Gly Val Arg Asn Ser Ser Ser
 2225 2230 2235 2240
 Ser Thr Ser Pro Val Ser Lys Lys Gly Pro Pro Leu Lys Thr Pro Ala
 2245 2250 2255
 Ser Lys Ser Pro Ser Glu Gly Gln Thr Ala Thr Thr Ser Pro Arg Gly
 2260 2265 2270
 Ala Lys Pro Ser Val Lys Ser Glu Leu Ser Pro Val Ala Arg Gln Thr
 2275 2280 2285
 Ser Gln Ile Gly Gly Ser Ser Lys Ala Pro Ser Arg Ser Gly Ser Arg
 2290 2295 2300
 Asp Ser Thr Pro Ser Arg Pro Ala Gln Gln Pro Leu Ser Arg Pro Ile
 2305 2310 2315 2320
 Gln Ser Pro Gly Arg Asn Ser Ile Ser Pro Gly Arg Asn Gly Ile Ser
 2325 2330 2335
 Pro Pro Asn Lys Leu Ser Gln Leu Pro Arg Thr Ser Ser Pro Ser Thr
 2340 2345 2350
 Ala Ser Thr Lys Ser Ser Gly Ser Gly Lys Met Ser Tyr Thr Ser Pro
 2355 2360 2365
 Gly Arg Gln Met Ser Gln Gln Asn Leu Thr Lys Gln Thr Gly Leu Ser
 2370 2375 2380
 Lys Asn Ala Ser Ser Ile Pro Arg Ser Glu Ser Ala Ser Lys Gly Leu
 2385 2390 2395 2400
 Asn Gln Met Asn Asn Gly Asn Gly Ala Asn Lys Lys Val Glu Leu Ser
 2405 2410 2415
 Arg Met Ser Ser Thr Lys Ser Ser Gly Ser Glu Ser Asp Arg Ser Glu
 2420 2425 2430

Arg Pro Val Leu Val Arg Gln Ser Thr Phe Ile Lys Glu Ala Pro Ser
 2435 2440 2445
 Pro Thr Leu Arg Arg Lys Leu Glu Glu Ser Ala Ser Phe Glu Ser Leu
 2450 2455 2460
 Ser Pro Ser Ser Arg Pro Ala Ser Pro Thr Arg Ser Gln Ala Gln Thr
 2465 2470 2475 2480
 Pro Val Leu Ser Pro Ser Leu Pro Asp Met Ser Leu Ser Thr His Ser
 2485 2490 2495
 Ser Val Gln Ala Gly Gly Trp Arg Lys Leu Pro Pro Asn Leu Ser Pro
 2500 2505 2510
 Thr Ile Glu Tyr Asn Asp Gly Arg Pro Ala Lys Arg His Asp Ile Ala
 2515 2520 2525
 Arg Ser His Ser Glu Ser Pro Ser Arg Leu Pro Ile Asn Arg Ser Gly
 2530 2535 2540
 Thr Trp Lys Arg Glu His Ser Lys His Ser Ser Ser Leu Pro Arg Val
 2545 2550 2555 2560
 Ser Thr Trp Arg Arg Thr Gly Ser Ser Ser Ser Ile Leu Ser Ala Ser
 2565 2570 2575
 Ser Glu Ser Ser Glu Lys Ala Lys Ser Glu Asp Glu Lys His Val Asn
 2580 2585 2590
 Ser Ile Ser Gly Thr Lys Gln Ser Lys Glu Asn Gln Val Ser Ala Lys
 2595 2600 2605
 Gly Thr Trp Arg Lys Ile Lys Glu Asn Glu Phe Ser Pro Thr Asn Ser
 2610 2615 2620
 Thr Ser Gln Thr Val Ser Ser Gly Ala Thr Asn Gly Ala Glu Ser Lys
 2625 2630 2635 2640
 Thr Leu Ile Tyr Gln Met Ala Pro Ala Val Ser Lys Thr Glu Asp Val
 2645 2650 2655
 Trp Val Arg Ile Glu Asp Cys Pro Ile Asn Asn Pro Arg Ser Gly Arg
 2660 2665 2670
 Ser Pro Thr Gly Asn Thr Pro Pro Val Ile Asp Ser Val Ser Glu Lys
 2675 2680 2685
 Ala Asn Pro Asn Ile Lys Asp Ser Lys Asp Asn Gln Ala Lys Gln Asn
 2690 2695 2700
 Val Gly Asn Gly Ser Val Pro Met Arg Thr Val Gly Leu Glu Asn Arg
 2705 2710 2715 2720
 Leu Asn Ser Phe Ile Gln Val Asp Ala Pro Asp Gln Lys Gly Thr Glu
 2725 2730 2735
 Ile Lys Pro Gly Gln Asn Asn Pro Val Pro Val Ser Glu Thr Asn Glu
 2740 2745 2750
 Ser Ser Ile Val Glu Arg Thr Pro Phe Ser Ser Ser Ser Ser Ser Lys
 2755 2760 2765

His Ser Ser Pro Ser Gly Thr Val Ala Ala Arg Val Thr Pro Phe Asn
 2770 2775 2780
 Tyr Asn Pro Ser Pro Arg Lys Ser Ser Ala Asp Ser Thr Ser Ala Arg
 2785 2790 2795 2800
 Pro Ser Gln Ile Pro Thr Pro Val Asn Asn Asn Thr Lys Lys Arg Asp
 2805 2810 2815
 Ser Lys Thr Asp Ser Thr Glu Ser Ser Gly Thr Gln Ser Pro Lys Arg
 2820 2825 2830
 His Ser Gly Ser Tyr Leu Val Thr Ser Val
 2835 2840

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:
 (B) CLONE: ral2(yeast)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Leu Thr Gly Ala Lys Gly Leu Gln Leu Arg Ala Leu Arg Arg Ile Ala
 1 5 10 15
 Arg Ile Glu Gln Gly Gly Thr Ala Ile Ser Pro Thr Ser Pro Leu
 20 25 30

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
 (B) CLONE: m3(MACHR)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Leu Tyr Trp Arg Ile Tyr Lys Glu Thr Glu Lys Arg Thr Lys Glu Leu
 1 5 10 15

Ala Gly Leu Gln Ala Ser Gly Thr Glu Ala Glu Thr Glu
 20 25

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:
 (B) CLONE: MCC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Leu Tyr Pro Asn Leu Ala Glu Glu Arg Ser Arg Trp Glu Lys Glu Leu
 1 5 10 15
 Ala Gly Leu Arg Glu Glu Asn Glu Ser Leu Thr Ala Met
 20 25

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 40 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTATCAAGAC TGTGAC TTT AATTGTAGTT TATCCATTT

40

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 40 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:12:
TTTAGAATTT CATGTTAATA TATTGTCTT TTTTAACAG

40

(2) INFORMATION FOR SEQ ID NO:13:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:13:
GTACATTTTA AAAAGGTGTT TTAAATAAT TTTTAAGCT

40

(2) INFORMATION FOR SEQ ID NO:14:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:14:
AAGCAATTGT TGTATAAAAA CTGTTTCTA TTTTATTAG

40

(2) INFORMATION FOR SEQ ID NO:15:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:15:
GTAACITTTT TICATATAGT AACATTGCC TTGTGTACTC

40

(2) INFORMATION FOR SEQ ID NO:16:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:16:

NNNNNNNNNN NNNGTCCCTT TTTTAAAAA AAAAAAATAG

40

(2) INFORMATION FOR SEQ ID NO:17:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GTAAGTAACT TGGCAGTACA ACTTATTGA AACTTTAATA

40

(2) INFORMATION FOR SEQ ID NO:18:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:18:

ATACAAGATA TTGATACTTT TTTATTATT GTGCTTTTAG

40

(2) INFORMATION FOR SEQ ID NO:19:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GTAAGTTACT TGTTCGTAAG TGATAAAACA CYGAAGACCT

40

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AATAAAACA TAACTAATTA GGTTCCTTGT TTTATTTTAG

40

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GTTAGTAAAT TSCCTTTTTT GTTGTGGGT ATAAAAATAG

40

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:22:
ACCATTTTTC CATGTACTGA TGTTAACTCC ATCTTAACAG

40

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:23:
GTAAATAAAT TATTTTATCA TATTTTITAA AATTATTTAA.

40

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 64 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:24:
CATGATGTTA TCTGTATTTA CCTATAGTCT AAATTATACC ATCTATAATG TGCTTAATTT
TTAG

60

64

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:25:
GTAACAGAAG ATTACAAACC CTGCTCACTA ATGCCATGAC TACTTIGCTA AG

52

(2) INFORMATION FOR SEQ ID NO:26:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 46 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CGATATTAAA CTCTAATT TCTTTCTAAA CTCATTTCG CACAC

46

(2) INFORMATION FOR SEQ ID NO:27:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GTATGTTCTC TATAGTGTAC ATCTAGTGC ATGTTTCAA

40

(2) INFORMATION FOR SEQ ID NO:28:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CATCATTGCT CTCAAATAA CAAACCTTA TGGTTTATGT TCATTTTATT TTTCAG

56

(2) INFORMATION FOR SEQ ID NO:29:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (11) MOLECULE TYPE: cDNA
(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:29:
GTAAGACAAA AATGTTTTT AATGACATAG ACAATTACTG GTC
(2) INFORMATION FOR SEQ ID NO:30:

43

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (11) MOLECULE TYPE: cDNA
(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:30:
TTAGATGATT GTCTTTTTC TCTGCCCTT TTAAATTAG
(2) INFORMATION FOR SEQ ID NO:31:

40

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (11) MOLECULE TYPE: cDNA
(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:31:
GTATGTTTTT ATAACATGTA TTTCTTAAGA TAGCTCAGGT ATCA
(2) INFORMATION FOR SEQ ID NO:32:

44

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (11) MOLECULE TYPE: cDNA
(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:32: 54
GCTTCGCTTC AACTTGNCTT TTAAATGATC CTCTATTCTC TATTAAATT ACAG

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 65 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:33: 60
GTACTATTA GAATTGACC TGTTTTCTT TTTCTCTT TCTTTGACC CAGGGTCTCA
CTCTG 65

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:34: 52
GCAACTACTA TGATTITATG TATAAATTAA TCTAAAATTG ATTAATTCC AC

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:35: 42
GTACCTTTGA AACATTAG TACTATAATA TGAATTTCAT GT

(2) INFORMATION FOR SEQ ID NO:36:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CCAACTCMAA TTAGATGACC CATATTCAGA AACTTACTAC

40

(2) INFORMATION FOR SEQ ID NO:37:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GTATATATAG ACTTTTATAT TACTTTTAAA GTACAGAATT CATACTCTCA AAAA

54

(2) INFORMATION FOR SEQ ID NO:38:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

ATTGTGACCT TAATTTTGTG ATCTCTTGAT TTTTATTCA G

41

(2) INFORMATION FOR SEQ ID NO:39:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (11) MOLECULE TYPE: cDNA
- (v1) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:39:

TCCCCCGCCTG CCGCTCTC

18

(2) INFORMATION FOR SEQ ID NO:40:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: cDNA
- (v1) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GCACCGCGCCG CTCCTCGT

18

(2) INFORMATION FOR SEQ ID NO:41:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: cDNA
- (v1) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GTGAACGGCT CTCATGCTGC

20

(2) INFORMATION FOR SEQ ID NO:42:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: cDNA
- (v1) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:42:

ACGTGCGGGG AGGAATGGA

19

(2) INFORMATION FOR SEQ ID NO:43:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:43:

ATGATATCTT ACCAATGAT ATAC

24

(2) INFORMATION FOR SEQ ID NO:44:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:44:

TTATTCCTAC TTCTTCIATA CAG

23

(2) INFORMATION FOR SEQ ID NO:45:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:45:

TACCCATGCT GGCTCTTTT C

21

(2) INFORMATION FOR SEQ ID NO:46:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TGCGGCCATC TTGTTCTCTGA

26

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

ACATTAGCCA CAAAGCTTGC AA

22

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

ATCAAGCTCC AGTAAGAAGC TA

22

(2) INFORMATION FOR SEQ ID NO:49:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (11) MOLECULE TYPE: cDNA
(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:49:

19

TGCGGCTCCT GGCTTGTTG

(2) INFORMATION FOR SEQ ID NO:50:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (11) MOLECULE TYPE: cDNA
(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:50:

20

GCCCCCTTCTT TTCTGAGGAC

(2) INFORMATION FOR SEQ ID NO:51:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (11) MOLECULE TYPE: cDNA
(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:51:

21

TTTTCTCCTG CCTCTTACTG C

(2) INFORMATION FOR SEQ ID NO:52:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (11) MOLECULE TYPE: cDNA
(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:52:
ATGACACCCC CCATTCCCTC

20

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:53:
CCACTTAAAG CACATATATT TAGT

24

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:54:
GTATCGAAAA TAGTGAAGAA CC

22

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:55:
TTCTTAAGTC CTCGTTTTCT TTTC

24

(2) INFORMATION FOR SEQ ID NO:56:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

TTTAGAACCT TTTTGTGTT CTG

23

(2) INFORMATION FOR SEQ ID NO:57:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CTCAGATTAT AACTAAGCC TAAC

24

(2) INFORMATION FOR SEQ ID NO:58:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

CATGCTCTT ACAGTAGTAC CA

22

(2) INFORMATION FOR SEQ ID NO:59:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

AGGTCCAAGG GTAGCCAAGG

20

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

TAAAAATGGA TAACTACAA TTAAAG

27

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

AAATACAGAA TCATGCTTG AAGT

24

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:62: 23
ACACCTAAAG ATGACAAATT CAG
- (2) INFORMATION FOR SEQ ID NO:63:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:63: 24
TAACTTAGAT AGCAGTAATT TCCC
- (2) INFORMATION FOR SEQ ID NO:64:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:64: 23
ACAATAAACT GGACTACACA AGC
- (2) INFORMATION FOR SEQ ID NO:65:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:65: 23
ATAGGTCATT GCTTCTTGCT CAT

(2) INFORMATION FOR SEQ ID NO:66:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

TCAATTTTAA TGGATTACCT AGGT

24

(2) INFORMATION FOR SEQ ID NO:67:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

CTTTTTTTGC TTTTACTGAT TAACG

25

(2) INFORMATION FOR SEQ ID NO:68:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

TCTAATTGAT TTTATTCCTA ATAGCTC

27

(2) INFORMATION FOR SEQ ID NO:69:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:69:

GGTAGCCATA GTATGATTAT TTCT

24

(2) INFORMATION FOR SEQ ID NO:70:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:70:

CTACCTATTT TTATACCCAC AAAC

24

(2) INFORMATION FOR SEQ ID NO:71:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:71:

AAGAAAGCCT ACACCTTTT TGC

23

(2) INFORMATION FOR SEQ ID NO:72:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

GATCATTCTT AGAACCATCT TGC

23

(2) INFORMATION FOR SEQ ID NO:73:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

ACCTATAGTC TAAATTATAC CATC

24

(2) INFORMATION FOR SEQ ID NO:74:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

GTCAATGGCAT TAGTGACCAG

20

(2) INFORMATION FOR SEQ ID NO:75:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

AGTCGTAATT TTGTTTCTAA ACTC

24

(2) INFORMATION FOR SEQ ID NO:76:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

TGAAGGACTC GGATTTCACG C

21

(2) INFORMATION FOR SEQ ID NO:77:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

TCATTCACCTC ACAGCCTCAT GAC

23

(2) INFORMATION FOR SEQ ID NO:78:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

GCTTTCAGAAC ATGCACTACG AT

22

(2) INFORMATION FOR SEQ ID NO:79:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA
(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:79:

AAACATCATT GCTCTTCAAA TAAC

24

(2) INFORMATION FOR SEQ ID NO:80:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA
(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:80:

TACCATGATT TAAAAATCCA CCAG

24

(2) INFORMATION FOR SEQ ID NO:81:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA
(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:81:

GATGATTGTC TTTTCTCTT TGC

23

(2) INFORMATION FOR SEQ ID NO:82:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA
(v1) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

CTGAGGTATC TTAACAAATA CATC

24

(2) INFORMATION FOR SEQ ID NO:83:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

TTTAAATGA TCCTCTATTC TGTAT

25

(2) INFORMATION FOR SEQ ID NO:84:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

ACAGAGTCAG ACCCTGCCTC AAAG

24

(2) INFORMATION FOR SEQ ID NO:85:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

TTTCTATTCT TACTGCTAGC ATT

23

(2) INFORMATION FOR SEQ ID NO:86:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

ATACACAGGT AAGAAATTAG GA

22

(2) INFORMATION FOR SEQ ID NO:87:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

TAGATGACCC ATATTCTGTT TC

22

(2) INFORMATION FOR SEQ ID NO:88:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

CAATTAGGTC TTTTGGAGAG TA

22

(2) INFORMATION FOR SEQ ID NO:89:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: cDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

GTTACTGCAT ACACATTGTC AC

22

(2) INFORMATION FOR SEQ ID NO:90:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

OCTTTTCTT TCCTAACATC AAG

23

(2) INFORMATION FOR SEQ ID NO:91:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

TCTCCACAG GTAATCTCC C

21

(2) INFORMATION FOR SEQ ID NO:92:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

CCTAGAACTG AATCGGGTAC G

21

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

CAGGACAAAA TAATCCTGTC CC

22

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

ATTTTCTTAG TTTCATTCTT CCTC

24

MICROORGANISMS	
Optional Sheet in connection with the microorganism referred to on page... <u>22</u> ..., line <u>23</u> ... of the description.	
A. IDENTIFICATION OF DEPOSIT	
Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution	
NATIONAL COLLECTION OF INDUSTRIAL AND MARINE BACTERIA (NCIMB)	
Address of depository institution (including postal code and country)	
23 St. Machar Drive Aberdeen AB2 1RY, Scotland United Kingdom	
Date of deposit	Accession Number
17 December 1990	NCIMB 40353
B. ADDITIONAL INDICATIONS (None given if not applicable). This information is continued on a separate attached sheet <input type="checkbox"/>	
Saccharomyces cerevisiae SC/37HG4	
C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (If the indications are not for all designated States)	
D. SEPARATE FURNISHING OF INDICATIONS (None given if not so required)	
The indications below have not been submitted to the International Bureau later (Specify the general nature of the indications and Accession Number of Deposit)	
E. <input checked="" type="checkbox"/> This sheet was received with the international application when filed or is attached to the received sheet	
<i>Shirley Harris</i> (Authorized Officer)	
<input type="checkbox"/> The date of receipt (from the applicant) by the International Bureau is	
... .. (Authorized Officer)	

CLAIMS

1. A method of diagnosing or prognosing a neoplastic tissue of a human, comprising:
detecting somatic alteration of wild-type APC gene coding sequences or their expression products in a tumor tissue isolated from a human, said alteration indicating neoplasia of the tissue.
2. The method of claim 1 wherein the expression products are mRNA molecules.
3. The method of claim 2 wherein the alteration of wild-type APC mRNA is detected by hybridization of mRNA from said tissue to an APC gene probe.
4. The method of claim 1 wherein alteration of wild-type APC gene coding sequences is detected by observing shifts in electrophoretic mobility of single-stranded DNA on non-denaturing polyacrylamide gels.
5. The method of claim 1 wherein alteration of wild-type APC gene coding sequences is detected by hybridization of an APC gene coding sequence probe to genomic DNA isolated from said tissue.
6. The method of claim 5 further comprising:
subjecting genomic DNA isolated from a non-neoplastic tissue of the human to Southern hybridization with the APC gene coding sequence probe; and
comparing the hybridizations of the APC gene probe to said tumor and non-neoplastic tissues.
7. The method of claim 5 wherein the APC gene probe detects a restriction fragment length polymorphism.
8. The method of claim 1 wherein the alteration of wild-type APC gene coding sequences is detected by determining the sequence of all or part of an APC gene in said tissue using a polymerase chain reaction, deviations in the APC sequence determined from that of the sequence shown in Figure 7 (SEQ ID NO.: 1) suggesting neoplasia.
9. The method of claim 1 wherein the alteration of wild-type APC gene coding sequences is detected by identifying a mismatch between molecules (1) an APC gene or APC mRNA isolated from said tissue and (2) a nucleic acid probe complementary to the human wild-

type APC gene coding sequence, when molecules (1) and (2) are hybridized to each other to form a duplex.

10. The method of claim 5 wherein the APC gene probe hybridizes to an exon selected from the group consisting of: (1) nucleotides 822 to 930; and (2) nucleotides 931 to 1309; (3) nucleotides 1406 to 1545; and (4) nucleotides 1956 to 2256.

11. The method of claim 1 wherein the alteration of wild-type APC gene coding sequences is detected by amplification of APC gene sequences in said tissue and hybridization of the amplified APC sequences to nucleic acid probes which comprise APC sequences.

12. The method of claim 1 wherein the alteration of wild-type APC gene coding sequences is detected by molecular cloning of the APC genes in said tissue and sequencing all or part of the cloned APC gene.

13. The method of claim 1 wherein the detection of alteration of wild-type APC gene coding sequences comprises screening for a deletion mutation.

14. The method of claim 1 wherein the detection of alteration of wild-type APC gene coding sequences comprises screening for a point mutation.

15. The method of claim 1 wherein the detection of alteration of wild-type APC gene coding sequences comprises screening for an insertion mutation.

16. The method of claim 1 wherein the tumor tissue is a colorectal tissue.

17. The method of claim 6 wherein the non-neoplastic tissue isolated from a human is from colonic mucosa.

18. The method of claim 1 wherein the expression products are protein molecules.

19. The method of claim 18 wherein the alteration of wild-type APC protein is detected by immunoblotting.

20. The method of claim 18 wherein the alteration of wild-type APC protein is detected by immunocytochemistry.

21. The method of claim 18 wherein the alteration of wild-type APC protein is detected by assaying for binding interactions between APC protein of said tumor tissue and a second cellular protein.

22. The method of claim 21 wherein the second cellular protein is selected from the group consisting of MCC protein, wild-type APC protein, and a G protein.

23. The method of claim 18 wherein the alteration of wild-type APC protein is detected by assaying for phospholipid metabolites.

24. A method of supplying wild-type APC gene function to a cell which has lost said function by virtue of a mutation in an APC gene, comprising:

introducing a wild-type APC gene into a cell which has lost said gene function such that said wild-type APC gene is expressed in the cell.

25. The method of claim 24 wherein the wild-type APC gene introduced recombines with the endogenous mutant APC gene present in the cell by a double recombination event to correct the APC gene mutation.

26. A method of supplying wild-type APC gene function to a cell which has altered APC function by virtue of a mutation in an APC gene, comprising:

introducing a portion of a wild-type APC gene into a cell which has lost said gene function such that said portion is expressed in the cell, said portion encoding a part of the APC protein which is required for non-neoplastic growth of said cell.

27. A method of supplying wild-type APC gene function to a cell which has altered APC function by virtue of a mutation in an APC gene, comprising:

applying human wild-type APC protein to a cell which has lost wild-type APC function.

28. A method of supplying wild-type APC gene function to a cell which has altered APC gene function by virtue of a mutation in an APC gene, comprising:

introducing into the cell a molecule which mimics the function of wild-type APC protein.

29. A pair of single stranded DNA primers for determination of a nucleotide sequence of an APC gene by polymerase chain reaction, the sequence of said primers being derived from chromosome 5q band 21, wherein the use of said primers in a polymerase chain reaction results in synthesis of DNA having all or part of the sequence shown in Figure 7.

30. The primers of claim 29 which have restriction enzyme sites at each 5' end.

31. The pair of primers of claim 29 having sequences corresponding to APC introns.

32. A nucleic acid probe complementary to human wild-type APC gene coding sequences.

33. The nucleic acid probe of claim 31 which hybridizes to an exon selected from the group consisting of: (1) nucleotides 822 to 930; and (2) nucleotides 931 to 1309; (3) nucleotides 1406 to 1545; (4) nucleotides 1956 to 2256.

34. A kit for detecting alteration of wild-type APC genes comprising a battery of nucleic acid probes which in the aggregate hybridize to all nucleotides of the APC gene coding sequences.

35. A method of detecting the presence of a neoplastic tissue in a human, comprising:

detecting in a body sample isolated from a human alteration of a wild-type APC gene coding sequence or wild-type APC expression product, said alteration indicating the presence of a neoplastic tissue in the human.

36. The method of claim 35 wherein said body sample is selected from the group consisting of serum, stool, urine and sputum.

37. A method of detecting genetic predisposition to cancer, including familial adenomatous polyposis (FAP) and Gardner's Syndrome (GS), in a human comprising:

detecting a germline alteration of wild-type APC gene coding sequences or their expression products in a human sample

selected from the group consisting of blood and fetal tissue, said alteration indicating predisposition to cancer.

38. The method of claim 37 wherein the expression products are mRNA molecules.

39. The method of claim 38 wherein the alteration of wild-type APC mRNA is detected by hybridization of mRNA from said tissue to an APC gene probe.

40. The method of claim 37 wherein alteration of wild-type APC gene coding sequences is detected by observing shifts in electrophoretic mobility of single-stranded DNA on non-denaturing polyacrylamide gels.

41. The method of claim 37 wherein alteration of wild-type APC gene coding sequences is detected by hybridization of an APC gene coding sequence probe to genomic DNA isolated from said tissue.

42. The method of claim 41 wherein the APC gene coding sequence probe detects a restriction fragment length polymorphism.

43. The method of claim 37 wherein the alteration of wild-type APC gene coding sequences is detected by determining the sequence of all or part of an APC gene in said tissue using a polymerase chain reaction, deviations in the APC sequence determined from the sequence of Figure 7 suggesting predisposition to cancer.

44. The method of claim 37 wherein the alteration of wild-type APC gene coding sequences is detected by identifying a mismatch between molecules (1) an APC gene or APC mRNA isolated from said tissue and (2) a nucleic acid probe complementary to the human wild-type APC gene coding sequence, when molecules (1) and (2) are hybridized to each other to form a duplex.

45. The method of claim 41 wherein the APC gene probe hybridizes to an exon selected from the group consisting of: (1) nucleotides 822 to 930; and (2) nucleotides 931 to 1309; (3) nucleotides 1406 to 1545 and (4) nucleotides 1956 to 2256.

46. The method of claim 37 wherein the alteration of wild-type APC gene coding sequences is detected by amplification of APC gene sequences in said tissue and hybridization of the amplified APC

sequences to nucleic acid probes which comprise APC gene coding sequences.

47. The method of claim 37 wherein the alteration of wild-type APC gene coding sequences is detected by molecular cloning of the APC genes in said tissue and sequencing all or part of the cloned APC gene.

48. The method of claim 37 wherein the detection of alteration of wild-type APC gene coding sequences comprises screening for a deletion mutation.

49. The method of claim 37 wherein the detection of alteration of wild-type APC gene coding sequences comprises screening for a point mutation.

50. The method of claim 37 wherein the detection of alteration of wild-type APC gene coding sequences comprises screening for an insertion mutation.

51. The method of claim 37 wherein the expression products are protein molecules.

52. The method of claim 51 wherein the alteration of wild-type APC protein is detected by immunoblotting.

53. The method of claim 51 wherein the alteration of wild-type APC protein is detected by immunocytochemistry.

54. The method of claim 51 wherein the alteration of wild-type APC protein is detected by assaying for binding interactions between APC protein isolated from said tissue and a second cellular protein.

55. The method of claim 54 wherein the second cellular protein is selected from the group consisting of MCC protein, wild-type APC protein and a G protein.

56. A method of screening for genetic predisposition to cancer, including familial adenomatous polyposis (FAP) and Gardner's Syndrome (GS), in a human comprising:

detecting among kindred persons the presence of a DNA polymorphism which is linked to a mutant APC allele in an individual having a genetic predisposition to cancer, said kindred being

genetically related to the individual, the presence of said polymorphism suggesting a predisposition to cancer.

57. A preparation of the human APC protein substantially free of other human proteins, the amino acid sequence of said protein corresponding to that shown in Figure 3 or 7 (SEQ ID NO: 1).

58. A preparation of antibodies immunoreactive with a human APC protein and not substantially immunoreactive with other human proteins.

59. A method of testing therapeutic agents for the ability to suppress a neoplastically transformed phenotype, comprising:

applying a test substance to a cultured epithelial cell which carries a mutation in an APC allele;

determining whether said test substance suppresses the neoplastically transformed phenotype of the cell.

60. The method of claim 59 wherein the cultured epithelial cell has been genetically engineered to carry the mutation in the APC allele.

61. A method of testing therapeutic agents for the ability to suppress neoplastic growth, comprising:

administering a test substance to an animal which carries a mutant APC allele in its genome;

determining whether said test substance prevents or suppresses the growth of tumors.

62. A transgenic animal which carries a mutant APC allele from a second animal species in its genome.

63. An animal which has been genetically engineered to contain an insertion mutation which disrupts an APC allele in its genome.

64. A cDNA molecule which encodes a protein having the amino acid sequence shown in Figure 3 or 7 (SEQ ID NO: 7 or 1).

65. An isolated DNA molecule which encodes a protein having the amino acid sequence shown in Figure 3 or 7 (SEQ ID NO: 7 or 1).

66. A yeast artificial chromosome which is known as 37HG4.

TABLE IIa
Germline mutations of the APC gene in FAP and GS Patients

<u>EXTRA-COLONIC</u> <u>PATIENT</u> <u>DISEASE</u>	<u>CODON</u>	<u>NUCLEOTIDE</u> <u>CHANGE</u>	<u>AMINO</u> <u>CHANGE</u>	<u>AGE</u>	<u>ACID</u>
93 Osteoma	279	TCA-→T <u>C</u> A	Ser-→Stop	29	Mandibular
34	301	CGA-→T <u>C</u> A	Arg-→Stop	46	None
34 Tumor	301	CGA-→T <u>C</u> A	Arg-→Stop	27	Desmoid
21 Osteoma	413	CGC-→T <u>C</u> C	Arg-→Cys	24	Mandibular
60 Osteoma	712	TCA-→T <u>C</u> A	Ser-→Stop	37	Mandibular
3746	243	CAGAG-→CAG	splice-junction		
3460	301	CGA-→T <u>C</u> A	Arg-→Stop		
3627	456	CTTTCa-→CTTCA	frameshift		
3712	500	T-→ <u>G</u>	Tyr-→Stop		

* The mutated nucleotides are underlined.

TABLE III

Somatic Mutations in Sporadic CRC Patients

<u>PATIENT</u>	<u>CODON¹</u>	<u>NUCLEOTIDE CHANGE</u>	<u>AMINO ACID CHANGE</u>
T35	MCC 12	GAG/ <u>gtag</u> → GAG/ <u>gtag</u>	(Splice Donor)
T16	MCC 145	<u>ctca</u> /GGA→ <u>gtca</u> /GGA	(Splice Acceptor)
T47	MCC 267	CGG→CTG	Arg→Leu
T81	MCC 490	TCG→TIG	Ser→Leu
T35	MCC 506	CGG→CAG	Arg→Gln
T91	MCC 698	GCT→GIT	Ala→Val
T34	APC 288	CCAGT→ <u>CCCAGCC</u> AGT	(Insertion)
T27	APC 331	CGA→TGA	Arg→Stop
T135	APC 437	CAA/ <u>gtta</u> →CAA/ <u>gtta</u>	(Splice Donor)
T201	APC 1338	CAG→TAG	Gln→Stop

For splice site mutations, the codon nearest to the mutation is listed

The underlined nucleotides were mutant; small case letters represent introns, large case letters represent exons

TABLE IV

Seven Different Versions of the 20-Amino Acid Repeat	
<u>Consensus: F * V E * T P * C F S R * S S L S S L S</u>	
1262:	Y C V E D T P I C F S R C S S L S S L S
1378:	H Y V Q E T P L M F S R C T S V S S L D
1492:	F A T E S T P D Q F S C S S S L S A L S
1643:	Y C V E G T P I N F S T A T S L S D L T
1848:	T P I E G T P Y C F S R N D S L S S L D
1953:	F A I E N T P V C P S H N S S L S S L S
2013:	F H V E D T P V C F S R N S S L S S L S

Numbers denote the first amino acid of each repeat. The consensus sequence at the top reflects a majority amino acid at a given position.

A) TB1 AMINO ACID SEQUENCE

VAPVVVGSGR APRHPAPAAH HPRRPDGF DG LGYRGGARDE OGFGGAFPAR SFSTGSDLEH 60
WYTPPDIPG SRNLHWGEKS PPYGVPTTST PYEGPTEEPF SSGGGGSVOG QSSEQLNRFA 120
GFGIGLASLF TENVLAHPCI VLRROCQVNY HAQHYHLTPF TVINIMYSFN KTOGPRALMX 180
GMGSTFIVOG VTLGAEGEIS EFTPLPREVL HKWSPKOIGE HLLKSLTYV VAMPFYSASL 240
IETVQSEIIR DNTGILECVK EGIGRVIGMG VPHSKRLLPL LSLIFPTVLH GVLHYIISV 300
IQKFVLLILK RKTYNKSHLAE STSPVQSHLD AYFPELIANF AASLCSDVIL YPLETVLHRL 360
HIQGTBTIID NTDLGYEVL INTQYEGMRD CINTIROEEG YEGFYKGFGA VIIQYTLHAA 420
VLQITKIIYS TLLQ 434

B) TB2 AMINO ACID SEQUENCE

ELRRFDRFLH EKNCHTOLLA KLEAKTGVNR SFIALGVIGL VALYLVFGYG ASLLCNLIGF 60
GYPAYISIKA IESPNKEDDT QWLTYWVYVG VFSIAEFFSD IFLSWFPFY ILKCGFLLWC 120
KAPSPNGAE LLYKRIIRPF FLKHESQKDS VVKDLKDKAX ETADAITKEA KKATVNLLGE 180
EKKST 185

FIGURE 2

APC AMINO ACID SEQUENCE

MAAASYDQLL	KQVEALKMEN	SNLROELEDM	SNHLTKLETE	ASHMKVCLKQ	LOGSIEDEAM	60
ASSGQIDLL	RLKELNLDSS	NFPGVKLRSK	MSLRSYGSRE	GSVSSRSSEC	SPVPMGSFPR	120
RGFVNGSRES	TGYLEELEKE	RSLLLADLQK	EEKEKQWYYA	QLQNLTKRID	SLLTENFSLQ	180
TOMTRROLEY	EARQIRVAME	EOLGTCQDME	KRAQRRIARI	QOIEKDILRI	ROLLQSQATE	240
AERSSQNKHE	TGSHDAERON	EGOGVGEIMH	ATSGNGOGST	TRMDHETASY	LSSSSSTHSAP	300
RRLTSHLGTK	VENVYSLLSM	LGTHDKDOMS	RTLLAMSSSO	DSCISMROSG	CLPLLIQLLH	360
GNOKDSVLLG	NSRGSKEARA	RASAALHNII	HSQPDOKRGR	REIRVLHLL	QIRAYCETCW	420
EWQEAHEPGH	DDQKNMPAP	VEHQICPAVC	YLMKLSFDEE	HRHAMWELGG	LOAJAELLOV	480
DCEMYGLTND	HYSITLRYA	GMALTNLTFG	DVANKATLCS	MKGCMRALVA	OLKSESEDLO	540
QVIASVLRNL	SWRADVNSKK	TLREVGSVKA	LMECALEYKK	ESTLKSVLSA	LWNLSAHCTE	600
NKADICAVDG	ALAFVGTLT	YRSQNTLAI	IESGGGILRN	VSSLIATNED	HRQILRENNC	660
LOTLLQHLKS	HSLTIVSNAC	GTLMNLSARN	PKDQEAALWM	GAVSNLKLHI	HSKXHXIANG	720
SAALRNLMHA	NRPAXYKDH	INSPGSSLP	LHVRKQKALE	AELDAQHLSE	TFQNDQNLSP	780
KASHRSKORH	KQSLYGDYVF	DTNRHDONRS	DNFHTGNHTV	LSPYLNTTVL	PSSSSSRGSL	840
DSSRSEKDRS	LERERIGLG	NYHPATENPG	TSSKRGLOIS	TTAAQIAKVM	EEVSAIHTSO	900
EDRSSGSTTE	LHCVTDERNA	LRRSSAANTH	SNTYNFTKSE	NSNRTCSMPY	AKLEYKRSSH	960
OSLNSVSSSD	GYGKRGQMKP	SIESYSEDOE	SKFCSYGOYP	ADLAHKXISA	NHMDQNDGEL	1020
OTPINYSLKY	SDEQLNSGRQ	SPSQNERMAR	PKHIEDEIK	QSEORQSRNQ	STTYPVYTES	1080
TDDKHLKFQ	HFGQQECVSP	YRSRGANGSE	THRVGSNHGI	NQNVSSQLCO	EDDYEDDKPT	1140
MYSEYSEEE	QHEEEERPTH	YSIKYNEEKR	HVDQPIDYSL	KYATDIPSSQ	KQSFSSFSKSS	1200
SGQSSKTEHM	SSSSENTSTP	SSNAKQONQL	HPSSAQSRSG	QPKAATCKY	SSINQETIQT	1260
YCVEDTPICF	SRCSLSLSL	SAEDEIGCNO	TTQDPOSANT	LQIAEIKEKI	GTRSAEDPVS	1320
EYPAVSQHPR	TKSSRLOGSS	LSSSARHKA	VEFSSGAKSP	SKSGAOTPKS	PPEHYVQETP	1380
LHFSRCTSVS	SLDSFESRSI	ASSVQSEPCS	GHVSGIISPS	DLPDSPGQTH	PPSRSKTPPP	1440
PPQTAQTKRE	VPKKXAPTAE	KRESGPKQAA	VNAAVQRVQV	LPOADTLLHF	ATESTPDGFS	1500
CSSSLSALS	DEPFIQKQVE	LRIMPVQEN	DNGHETESQ	PKESHENQEK	EAEKTIDSEK	1560
DLLDDSDDD	IEILEECIIS	AMPTKSSRKA	KQPAQTASKL	PPPVARKPSQ	LPVYKLLPSQ	1620
NRLOPQKHVS	FTPGQONPRV	YCVEGTPINF	STATSLDOLT	IESPPNELAA	GEGVRGGAQS	1680
GEFEXROTIP	TEGRSTDEAQ	GGKTSSVTIP	ELDDNKAEEG	DILAECINSA	MPKKGSKHPF	1740
RVKXIMQVQ	QASASSSAPH	KHOLDGKQCK	PTSPVKPIPO	NTEYRTRVRK	NADSKHNLNA	1800
ERVFSQNKDS	KKQNLQKMSK	DFNDKLPNNE	DRVRGSFAFD	SPHHYTPIEG	TPYCFSRNDS	1860
LSSLDFODD	VDLREKAEAL	RKAKENKESE	AKYTSHTELT	SHQDSANKTO	AIKQPINRG	1920
QPKPILOKDS	TFPOSSKQIP	DRGAATDEKL	QNFATENTPV	CFSHNSSLSS	LSDIDQENHN	1980
KENEPKETE	PPDSQGEPSK	PQASGYAPKS	FHVEDTPVCF	SRNSSLSSLS	IDSEDOLLOE	2040
CISSAMPKKK	KPSRLKGONE	KHSRPMGGI	LGEDLTLDLK	DIQRPOSENG	LSPDSENFOW	2100
KAIQEGANSI	VSSLHQAAAA	ACLSROASSD	SUSTLSLKS	ISLGSPFHLT	PQEEKPFTS	2160
NKGPRIKPG	EKSTLETIXI	ESESKGIKGG	KQVYKSLITG	KVRSHSEISG	QMKQPLQANH	2220
PSISRGRTNI	HIPGVRNSSS	STSPVSKXGP	PLKTPASKSP	SEGQTATTSP	RGAXPSVKSE	2280
LSPVARQTSQ	IGGSSKAPSR	SGSRDSTPSR	PADQPLSRPI	QSPGRNSISP	GRNGISPPNK	2340
LSQLPRTSSP	STASTKSSGS	GKMSYTSRGR	QMSQNLTKQ	TGLSKNASSI	PRSESASKGL	2400
NQHNHNGGAN	KKVELSRMSS	TKSSGSESDR	SERPVLVROS	TFIKEAPSPT	LRRKLEESAS	2460
FESLSPSSRP	ASPTRSQOAT	PVLSPSLPDM	SLSTHSSVQA	GGWRKLPPML	SPTIEYNDGR	2520
PAKRHDZARS	HSESPSRLLPI	NRSGTWKREH	SKHSSSLPRV	STWRTGSSS	SILSASSESS	2580
EKAKEDEKX	VNSISGTKQS	KENQVSAGT	WRKIKENEF	PTNSTSQTYS	SGATNGAESK	2640
TLIYQMAPAV	SKTEDVWRI	EDCPINHPRS	GRSPTGNTTP	VIDSVEKAN	PHIKDSKQNO	2700
AKQNVGNQSV	PHRTVGLNR	LNSFIOVDAP	DQKGTETKPG	QNNPVPVSET	NESSIVERTP	2760
FSSSSSSKHS	SPSGTYAARY	TPFKYHPSPR	KSSADSTSR	PSQIPTPVHN	NTKKRDSKTD	2820
STESSGTQSP	KRHSGSYLVT	SV				2842

FIGURE 1

A

APC	203	LGTCQDHEKRAORRIARIQDIEKDILRIQL	233
		:: :	
RAL2	576	LTGAKGLQLRALRRIARIEQGGTAISPTSPL	606

B

APC	453	MKLSFDEEHRHANNELGGLOAIAELLOVD	481
		: : : : :	
M3 MACHR	249	LYWRIYKETEXRTKELAGLOASGTEAETE	277
		: . :	
MCC	220	LYPNLAERSRWEXELAGLREENESLTAM	248
		: : : :	
APC	453	MKLSFDEEHRHANNELGGLOAIAELLOVD	481

FIGURE 4

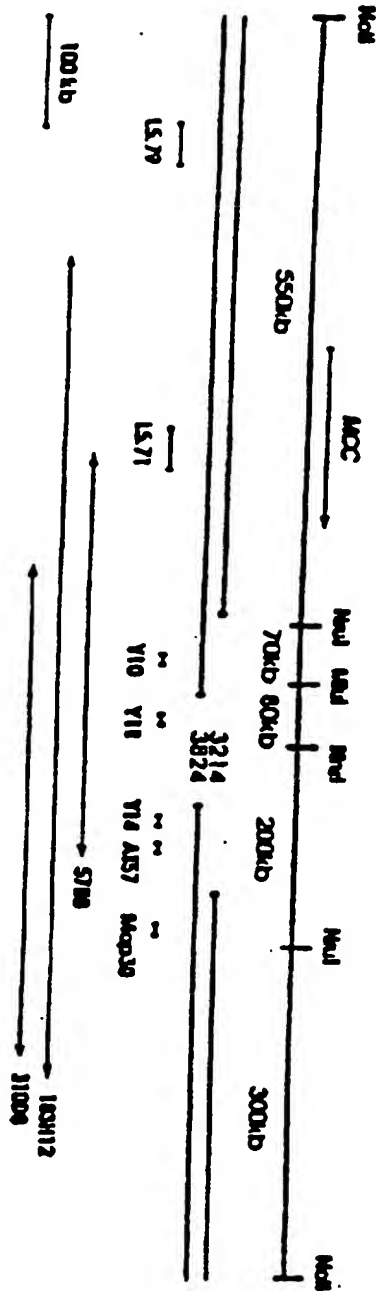


FIGURE 5

6/11

21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

[illegible]

Figure 6

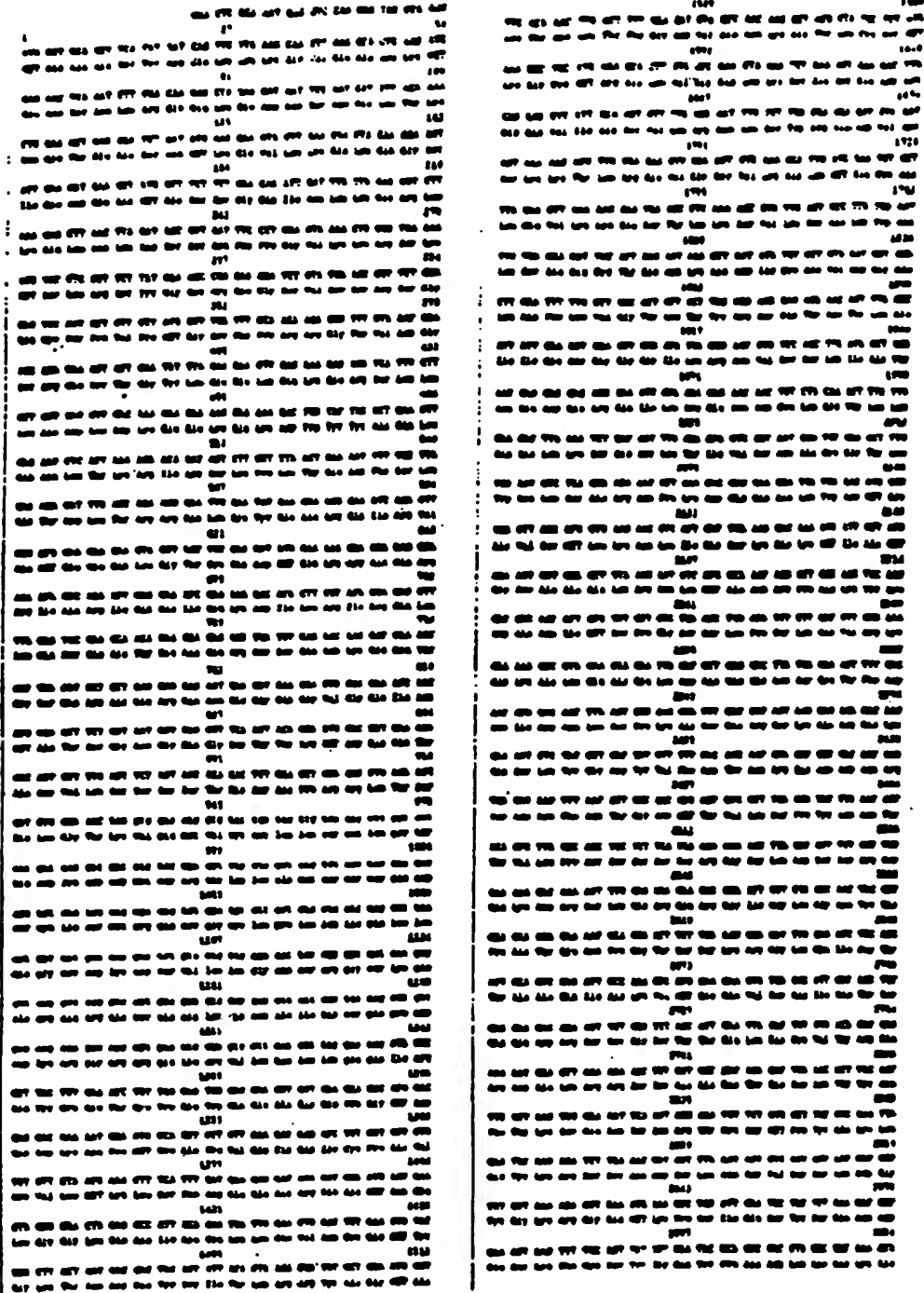


FIGURE 7

0000
0001
0002
0003
0004
0005
0006
0007
0008
0009
0010
0011
0012
0013
0014
0015
0016
0017
0018
0019
0020
0021
0022
0023
0024
0025
0026
0027
0028
0029
0030
0031
0032
0033
0034
0035
0036
0037
0038
0039
0040
0041
0042
0043
0044
0045
0046
0047
0048
0049
0050
0051
0052
0053
0054
0055
0056
0057
0058
0059
0060
0061
0062
0063
0064
0065
0066
0067
0068
0069
0070
0071
0072
0073
0074
0075
0076
0077
0078
0079
0080
0081
0082
0083
0084
0085
0086
0087
0088
0089
0090
0091
0092
0093
0094
0095
0096
0097
0098
0099
0100
0101
0102
0103
0104
0105
0106
0107
0108
0109
0110
0111
0112
0113
0114
0115
0116
0117
0118
0119
0120
0121
0122
0123
0124
0125
0126
0127
0128
0129
0130
0131
0132
0133
0134
0135
0136
0137
0138
0139
0140
0141
0142
0143
0144
0145
0146
0147
0148
0149
0150
0151
0152
0153
0154
0155
0156
0157
0158
0159
0160
0161
0162
0163
0164
0165
0166
0167
0168
0169
0170
0171
0172
0173
0174
0175
0176
0177
0178
0179
0180
0181
0182
0183
0184
0185
0186
0187
0188
0189
0190
0191
0192
0193
0194
0195
0196
0197
0198
0199
0200
0201
0202
0203
0204
0205
0206
0207
0208
0209
0210
0211
0212
0213
0214
0215
0216
0217
0218
0219
0220
0221
0222
0223
0224
0225
0226
0227
0228
0229
0230
0231
0232
0233
0234
0235
0236
0237
0238
0239
0240
0241
0242
0243
0244
0245
0246
0247
0248
0249
0250
0251
0252
0253
0254
0255
0256
0257
0258
0259
0260
0261
0262
0263
0264
0265
0266
0267
0268
0269
0270
0271
0272
0273
0274
0275
0276
0277
0278
0279
0280
0281
0282
0283
0284
0285
0286
0287
0288
0289
0290
0291
0292
0293
0294
0295
0296
0297
0298
0299
0300
0301
0302
0303
0304
0305
0306
0307
0308
0309
0310
0311
0312
0313
0314
0315
0316
0317
0318
0319
0320
0321
0322
0323
0324
0325
0326
0327
0328
0329
0330
0331
0332
0333
0334
0335
0336
0337
0338
0339
0340
0341
0342
0343
0344
0345
0346
0347
0348
0349
0350
0351
0352
0353
0354
0355
0356
0357
0358
0359
0360
0361
0362
0363
0364
0365
0366
0367
0368
0369
0370
0371
0372
0373
0374
0375
0376
0377
0378
0379
0380
0381
0382
0383
0384
0385
0386
0387
0388
0389
0390
0391
0392
0393
0394
0395
0396
0397
0398
0399
0400
0401
0402
0403
0404
0405
0406
0407
0408
0409
0410
0411
0412
0413
0414
0415
0416
0417
0418
0419
0420
0421
0422
0423
0424
0425
0426
0427
0428
0429
0430
0431
0432
0433
0434
0435
0436
0437
0438
0439
0440
0441
0442
0443
0444
0445
0446
0447
0448
0449
0450
0451
0452
0453
0454
0455
0456
0457
0458
0459
0460
0461
0462
0463
0464
0465
0466
0467
0468
0469
0470
0471
0472
0473
0474
0475
0476
0477
0478
0479
0480
0481
0482
0483
0484
0485
0486
0487
0488
0489
0490
0491
0492
0493
0494
0495
0496
0497
0498
0499
0500
0501
0502
0503
0504
0505
0506
0507
0508
0509
0510
0511
0512
0513
0514
0515
0516
0517
0518
0519
0520
0521
0522
0523
0524
0525
0526
0527
0528
0529
0530
0531
0532
0533
0534
0535
0536
0537
0538
0539
0540
0541
0542
0543
0544
0545
0546
0547
0548
0549
0550
0551
0552
0553
0554
0555
0556
0557
0558
0559
0560
0561
0562
0563
0564
0565
0566
0567
0568
0569
0570
0571
0572
0573
0574
0575
0576
0577
0578
0579
0580
0581
0582
0583
0584
0585
0586
0587
0588
0589
0590
0591
0592
0593
0594
0595
0596
0597
0598
0599
0600
0601
0602
0603
0604
0605
0606
0607
0608
0609
0610
0611
0612
0613
0614
0615
0616
0617
0618
0619
0620
0621
0622
0623
0624
0625
0626
0627
0628
0629
0630
0631
0632
0633
0634
0635
0636
0637
0638
0639
0640
0641
0642
0643
0644
0645
0646
0647
0648
0649
0650
0651
0652
0653
0654
0655
0656
0657
0658
0659
0660
0661
0662
0663
0664
0665
0666
0667
0668
0669
0670
0671
0672
0673
0674
0675
0676
0677
0678
0679
0680
0681
0682
0683
0684
0685
0686
0687
0688
0689
0690
0691
0692
0693
0694
0695
0696
0697
0698
0699
0700
0701
0702
0703
0704
0705
0706
0707
0708
0709
0710
0711
0712
0713
0714
0715
0716
0717
0718
0719
0720
0721
0722
0723
0724
0725
0726
0727
0728
0729
0730
0731
0732
0733
0734
0735
0736
0737
0738
0739
0740
0741
0742
0743
0744
0745
0746
0747
0748
0749
0750
0751
0752
0753
0754
0755
0756
0757
0758
0759
0760
0761
0762
0763
0764
0765
0766
0767
0768
0769
0770
0771
0772
0773
0774
0775
0776
0777
0778
0779
0780
0781
0782
0783
0784
0785
0786
0787
0788
0789
0790
0791
0792
0793
0794
0795
0796
0797
0798
0799
0800
0801
0802
0803
0804
0805
0806
0807
0808
0809
0810
0811
0812
0813
0814
0815
0816
0817
0818
0819
0820
0821
0822
0823
0824
0825
0826
0827
0828
0829
0830
0831
0832
0833
0834
0835
0836
0837
0838
0839
0840
0841
0842
0843
0844
0845
0846
0847
0848
0849
0850
0851
0852
0853
0854
0855
0856
0857
0858
0859
0860
0861
0862
0863
0864
0865
0866
0867
0868
0869
0870
0871
0872
0873
0874
0875
0876
0877
0878
0879
0880
0881
0882
0883
0884
0885
0886
0887
0888
0889
0890
0891
0892
0893
0894
0895
0896
0897
0898
0899
0900
0901
0902
0903
0904
0905
0906
0907
0908
0909
0910
0911
0912
0913
0914
0915
0916
0917
0918
0919
0920
0921
0922
0923
0924
0925
0926
0927
0928
0929
0930
0931
0932
0933
0934
0935
0936
0937
0938
0939
0940
0941
0942
0943
0944
0945
0946
0947
0948
0949
0950
0951
0952
0953
0954
0955
0956
0957
0958
0959
0960
0961
0962
0963
0964
0965
0966
0967
0968
0969
0970
0971
0972
0973
0974
0975
0976
0977
0978
0979
0980
0981
0982
0983
0984
0985
0986
0987
0988
0989
0990
0991
0992
0993
0994
0995
0996
0997
0998
0999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
204

000	001	002	003	004	005	006	007	008	009	010	011	012	013	014	015	016	017	018	019	020	021	022	023	024	025	026	027	028	029	030	031	032	033	034	035	036	037	038	039	040	041	042	043	044	045	046	047	048	049	050	051	052	053	054	055	056	057	058	059	060	061	062	063	064	065	066	067	068	069	070	071	072	073	074	075	076	077	078	079	080	081	082	083	084	085	086	087	088	089	090	091	092	093	094	095	096	097	098	099	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

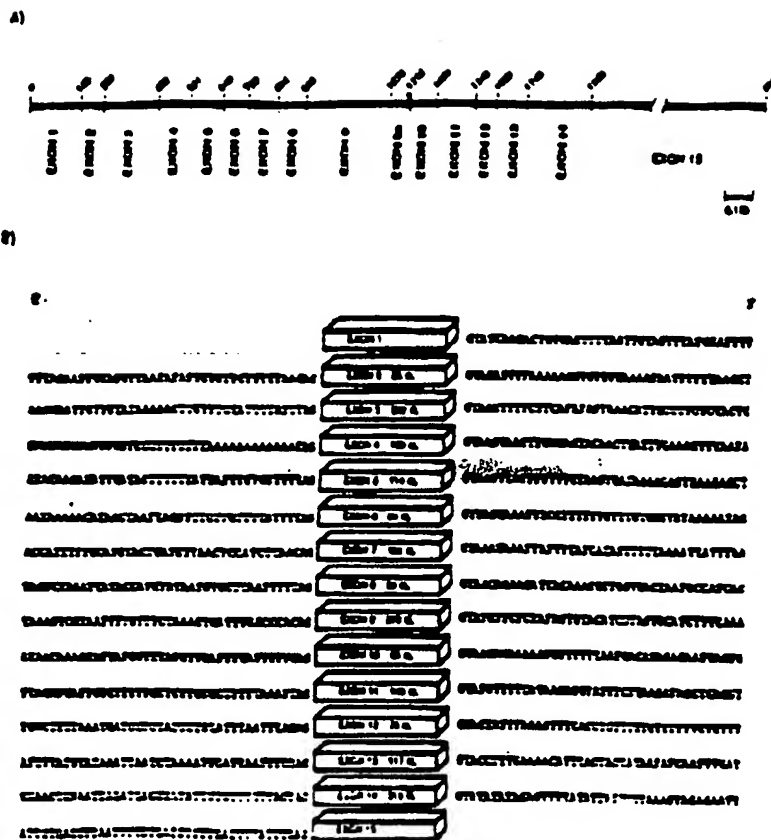


FIGURE 8

INTERNATIONAL SEARCH REPORT

PCT/US 92/00376

International Application No.

I. CLASSIFICATION OF SUBJECT MATTER of several classification systems apply, indicate only: According to International Patent Classification (IPC) or to both National Classification and IPC Int. Cl. 5 C12Q1/68; C12N15/12		
II. FIELDS SEARCHED Minimum Documentation Searched:		
Classification System	Classification Systems	
Int. Cl. 5	C12Q ; C07K ; G01N	
Documents Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched:		
III. DOCUMENTS CONSIDERED TO BE RELEVANT¹		
Category ²	Class of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
A	WO,A,8 901 481 (IMPERIAL CANCER RESEARCH TECHNOLOGY) 23 February 1989 see page 2, line 15 - page 6, line 12; claims	1
A	BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS. vol. 174, no. 1, 15 January 1991, DULUTH, MINNESOTA US pages 298 - 304; Y.HOSHINO ET AL.: 'Normal human chromosome 5, on which a familial adenomatous polyposis gene is located, has tumor suppressive activity' see abstract see page 302, line 1 - page 303, line 9	26
A	WO,A,9 005 180 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 17 May 1990 see page 7, line 25 - page 9, line 28; claims	26-28
-/-		
¹⁰ Special categories of cited documents: ¹⁰ ¹⁴ "A" documents defining the general state of the art which is not considered to be of particular relevance ¹⁵ "E" earlier document but published on or after the international filing date ¹⁶ "L" documents which may throw doubts on priority claim(s) or which is cited to establish the prior art date of another claim or other special reason (as specified) ¹⁷ "O" document relating to an oral disclosure, use, exhibition or other means ¹⁸ "P" document published prior to the international filing date but later than the priority date claimed ¹⁹ "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention ²⁰ "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step ²¹ "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art ²² "Z" document member of the same patent family		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search 01 JUNE 1992		Date of Mailing of this International Search Report 09 JUN 1992
International Searching Authority EUROPEAN PATENT OFFICE		Signature of Authorised Officer LUZZATTO E.R.

II. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Character of Document, with citations, where appropriate, of the relevant passages	Relevant to Claim No.
P,X	SCIENCE. vol. 253, 9 August 1991, LANCASTER, PA US pages 661 - 665; K.W.KINZLER ET AL.: 'Identification of FAP locus genes from chromosome 5q21' see the whole document ---	1,29,32, 64-66
P,X	SCIENCE. vol. 253, 9 August 1991, LANCASTER, PA US pages 665 - 669; I.NISHISHO ET AL.: 'Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients' see the whole document ---	1,37
T	SCIENCE. vol. 253, 9 August 1991, LANCASTER, PA US page 616; J.MARK: 'Gene identified for inherited cancer susceptibility' ---	

ANNEX TO THE INTERNATIONAL SEARCH REPORT
ON INTERNATIONAL PATENT APPLICATION NO. US 9200376
SA 57001

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The documents are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for those particulars which are merely given for the purpose of information. 01/06/92

Patent document cited in search report	Publication date	Patent family number(s)	Publication date
WO-A-8901481	23-02-89	EP-A- 0376968	11-07-90
		JP-T- 3503838	29-08-91
		US-A- 5098823	24-03-92
WO-A-9005180	17-05-90	AU-A- 4635089	28-05-90
		CA-A- 2001815	30-04-90
		EP-A- 0440744	14-08-91
		JP-T- 3505675	12-12-91

EPO FORM P479

For more details about this annex: see Official Journal of the European Patent Office, No. 12/82